

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

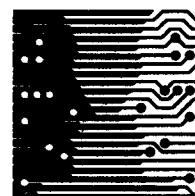
Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

**A Comparison of Statistical and Geometric Reconstruction Techniques:  
Guidelines for Correcting Fossil Hominin Crania**

**A DISSERTATION  
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE,  
FACULTY OF SCIENCE  
AT THE UNIVERSITY OF CAPE TOWN  
IN FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
MASTER OF SCIENCE**

**By  
Rudolph Neeser  
February 2007**

**Supervised by  
James Gain and Rebecca Rogers Ackermann**



© Copyright 2007  
by  
**Rudolph Neeser**

*For my family.  
And,  
ultimately,  
just for myself.*

“Welcome to Tokyo.”



# Abstract

The study of human evolution centres, to a large extent, around the study of fossil morphology, including the comparison and interpretation of these remains within the context of what is known about morphological variation within living species. However, many fossils suffer from environmentally caused damage (taphonomic distortion) which hinders any such interpretation: fossil material may be broken and fragmented while the weight and motion of overlaying sediments can cause their plastic distortion. To date, a number of studies have focused on the reconstruction of such taphonomically damaged specimens. These studies have used myriad approaches to reconstruction, including thin plate spline methods, mirroring, and regression-based approaches. The efficacy of these techniques remains to be demonstrated, and it is not clear how different parameters (e.g., sample sizes, landmark density, etc.) might effect their accuracy.

In order to partly address this issue, this thesis examines three techniques used in the *virtual reconstruction* of fossil remains by statistical or geometrical means: *mean substitution*, *thin plate spline warping* (TPS), and *multiple linear regression*. These methods are compared by reconstructing the same sample of individuals using each technique. Samples drawn from *Homo sapiens*, *Pan troglodytes*, *Gorilla gorilla*, and various hominin fossils are reconstructed by iteratively removing then estimating the landmarks. The testing determines the methods' behaviour in relation to the extant of landmark loss (i.e., amount of damage), reference sample sizes (this being the data used to guide the reconstructions), and the species of the population from which the reference samples are drawn (which may be different to the species of the damaged fossil).

Given a large enough reference sample, the regression-based method is shown to produce the most accurate reconstructions. Various parameters effect this: when using small reference samples drawn from a population of the same species as the damaged specimen, thin plate splines is the better method, but only as long as there is little damage. As the damage becomes severe (missing 30% of the landmarks, or more), mean substitution should be used instead: thin plate splines are shown to have a rapid error growth in relation to the amount of damage. When the species of the damaged specimen is unknown, or it is the only known individual of its species, the smallest reconstruction errors are obtained with a regression-based approach using a large reference sample drawn from a living species. Testing shows that reference sample size (combined with the use of multiple linear regression) is more important than morphological similarity between the reference individuals and the damaged specimen.

The main contribution of this work are recommendations to the researcher on which of the three methods to use, based on the amount of damage, number of reference individuals, and species of the reference individuals.



# Acknowledgements

Well, this is it; this is my masters. It's odd to see it finally finished, and it's more than usually difficult to look at the experience with only an objective eye. Subjectively, this has been one of the more difficult things that I've had to do. At its end the work became an uphill struggle, trying to balance research and a full time job on an emotional landscape that — if the reader would allow me to freely mix metaphors — could be described as quickly fraying. This, along with my original naïvety about what was required in doing honest research, was probably the biggest hurdle that I had to overcome during the length of my masters.

But now, when I hold these same pages that you, dear reader, are holding, I would like to think that I've learnt something about research. And if I have, there are some people that should be thanked for that. The most obvious — and most important besides — are James Gain and Rebecca Rogers Ackermann. I'm sure that as much as this work was my chore, I was one of theirs, and they quite gracefully helped me through it.

Because I *didn't* manage to live as a complete hermit for the last few years of my life, I have known various other people that have come and gone and that have effected both me and my masters. While most of this dissertation is concerned with my thesis, this section, at least, can talk about them.

A Masters in the *Collaborative Visual Computing Laboratory* of UCT's Computer Science department is not a solitary thing, and over the years the members of the lab — and the other labs in the department — have become close friends, almost an extended family of sorts. I'd like to thank some of them, and in no particular order they are: Dave Nunez, Ilda Ladeira, Shaun Nirenstein, Patrick Marais, Simon Perkins, Jakkaphan Tangkuampien, Cara Winterbottom, Hendranus Vermeulen, Carl Hultquist, Ilan Angel, Bryan Wong, Bruce Merry, Zayd Hendricks, Sarah Brown. It's a long list and I haven't named them all.

There are various people outside of the lab that need thanking:

- Prof. Heinz Rüther and Dr. Julian Smit (of UCT's Geomatics Department), for their assistance with the photogrammetric work that I had to perform.

- Stephany Potze (Transvaal museum, Northern Flagship Institution) and Dr. Kevin Kuykendall (at the time with the Anatomy Department of WITS) for graciously helping me in seeing and photographing various fossil specimens in the care of their respective institutions.
- Lal Adams, for performing casting work for me.
- Colin Trudox (at UCT's Psychology Department), for the use of the facial images in figure 3.
- James & Ingrid Gain, with assorted cats, for loaning me a house that I spent a lot of my life living in while working on my masters.

This work was funded by various people: the *National Research Foundation* and the *Paleo-Anthropological Scientific Trust*.

Due to the vagaries of fate — I forget why, and I'm sure Dave doesn't even remember at all, but it was probably a sordid trade of some kind — I should be thanking Dave one more time. Thank you Dave!

Having spoken subjectively, what can I say objectively about the work itself? Well, to that end I wrote this dissertation. I'll let it speak for itself, and hope that the reader finds some value in it.

# Notes for the Reader

## Terminology

Milliron et al. (2002) note that the Computer Science literature dealing with techniques for indirectly modifying the geometric shape of an object by changing the object's surrounding space tend to use the words *deformation* and *warp* — the terms that describe these techniques — interchangeably. Unfortunately, doing so in this dissertation could cause unnecessary confusion. Instead, the simple approach is taken that *deformation* refers to the *damage* of a fossil specimen that the techniques discussed here are used to correct. *Distortion* will be used synonymously. The techniques used to correct distortion of a fossil specimen via interpolation methods are *warps* and *warping* techniques. Unfortunately, there are some exceptions to this, but they are easy enough to remember: some notable warping algorithms bear the word “deformation” in their title, such as *Free Form Deformation* (Sederberg and Parry, 1986). These will be spoken of as warping techniques, although their names will continue to use the word “deformation”.

## Some Mathematical Notes

Generally, all matrices are given in upper case,  $A, B, C$ , etc. Rotation matrices are typically represented by  $R$ . Vectors are in lower case, as with  $\vec{a}$ ; translation vectors are usually notated as  $\vec{t}$ . Unless stated otherwise, or being obvious from context, all vectors themselves are column vectors. This means that vectors will be premultiplied by matrices:  $\vec{x}$  rotated by  $R$  would be the vector  $R\vec{x}$ .

Note that this thesis makes no use of projective geometry (i.e., homogenous coordinates) to unify the representation of rotations and translations.

Points are represented as vectors rather than as separate entities.

The matrix and vector transpose operators are represented as  $.^T$ . When components of a matrix are discussed, they are notated in lower case, with subscripted indices:  $e_{ij}$  is the component in row  $i$ , column  $j$  of matrix  $E$ . It is then said that  $E = (e_{ij})$ .

Sets are generally given as  $\{x_i\}_{i=1}^n$ , which indicates a set with values  $x_i$ , with  $1 \leq i \leq n$ . This will often be abbreviated as  $\{x_i\}$ . A constructor notation for sets is also used:  $\{\forall x | x < 5\}$  which is the set of all numbers less than five.  $\forall x$  and  $\exists x$  have their usual meanings: *for all*  $x$  and *there exists* an  $x$ , respectively. In this notation,  $\{x_i\}_{i=1}^n = \{\forall x_i | 1 \leq i \leq n\}$ .

$\in$  denotes set membership:  $s \in S$  is read as “ $s$ , an element of set  $S$ ”. The union of any two sets,  $A$  and  $B$ , is shown as  $A \cup B$ .  $A$  being a subset of  $B$  is notated as  $A \subset B$ .

The real numbers are represented as  $\mathbb{R}$ , and the set of all natural numbers (including zero) as,  $\mathbb{N}$ .  $\mathbb{N}^+$  and  $\mathbb{R}^+$  are the sets of all positive natural and real numbers, respectively.

$tr(X)$  is the trace of a matrix  $X$  (i.e., the sum of its diagonal elements). On occasion, when one variable must be shown as being proportional to another, the notation  $a \propto b$  — read as *variable a is proportional to b* — is used, meaning that there exists some real number  $x$  where  $a = xb$ . This generalises in the obvious way to vectors and matrices.

# Contents

<b>Abstract</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>Notes for the Reader</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The Fossil Record . . . . .	2
1.2 Taphonomic Distortion . . . . .	2
1.3 Motivation, Aims and Research Questions . . . . .	4
1.3.1 Motivation . . . . .	4
1.3.2 Aims and Research Questions . . . . .	7
1.4 Contributions and Results . . . . .	10
1.5 An Overview of the Dissertation . . . . .	11
<b>2 Background</b>	<b>13</b>
2.1 Fossilisation . . . . .	13
2.2 Taphonomy: “The Laws of Burial” . . . . .	14
2.3 Australopith Fossils in South Africa: Sites and Taphonomy . . . . .	15
2.4 Australopith Fossils in eastern Africa . . . . .	18
2.5 Australopith Fossils in central Africa . . . . .	18
2.6 Morphometrics: Basic Concepts . . . . .	19
2.7 Previous Reconstruction Examples . . . . .	22
2.7.1 Overview . . . . .	22

2.7.2	The Examples . . . . .	23
2.7.3	Various Observations and Problems . . . . .	25
<b>3</b>	<b>Data Acquisition</b>	<b>31</b>
3.1	Acquisition Techniques . . . . .	31
3.1.1	Contact Digitisers . . . . .	32
3.1.2	Diagnostic Radiology . . . . .	33
3.1.3	Laser Range Scanning . . . . .	34
3.1.4	Photogrammetry . . . . .	36
3.1.5	Data Forms . . . . .	37
3.2	Photogrammetric Outline . . . . .	38
3.2.1	The Coordinate System . . . . .	39
3.2.2	Photography . . . . .	41
3.2.3	Determination of External Orientation . . . . .	42
3.2.4	Locating Interest Points . . . . .	42
3.2.5	Epipolar Geometry . . . . .	43
<b>4</b>	<b>Data Processing and Display</b>	<b>47</b>
4.1	The Form of the Data . . . . .	47
4.2	Data Processing . . . . .	49
4.2.1	Processing of Volumetric Data . . . . .	51
4.2.2	Processing Surface Data . . . . .	53
4.3	Rendering . . . . .	58
4.3.1	Rendering Point Clouds . . . . .	59
4.4	Closing Remarks . . . . .	60
<b>5</b>	<b>The Correction Techniques</b>	<b>61</b>
5.1	Aspects of Statistical and Geometric Reconstruction . . . . .	61
5.1.1	Landmarks and Semilandmarks . . . . .	62
5.1.2	Sample Sizes . . . . .	63
5.1.3	Multidimensional Scaling . . . . .	65

5.1.4	Registration . . . . .	70
5.1.5	Centroid Size . . . . .	75
5.2	Reconstruction Techniques . . . . .	75
5.2.1	Mean Substitution . . . . .	76
5.2.2	Thin Plate Spline Warping . . . . .	78
5.2.3	Multiple Linear Regression . . . . .	82
5.2.4	Comments on the Methods . . . . .	87
5.2.5	Correcting Interlandmark Morphology . . . . .	87
<b>6</b>	<b>Tests and Results</b>	<b>89</b>
6.1	Materials and Methods . . . . .	89
6.1.1	Analysis I — The Mean substitution Method . . . . .	91
6.1.2	Analysis II — The Thin Plate Splines Method . . . . .	93
6.1.3	Analysis III — The Regression Method . . . . .	94
6.1.4	Analysis IV — Cross Over Point . . . . .	95
6.1.5	Analysis V — Cumulative Errors . . . . .	98
6.1.6	Analysis VI — Landmark Error Spread . . . . .	103
6.1.7	Analysis VII — Correcting Fossil Specimens . . . . .	104
<b>7</b>	<b>Discussion and Recommendation</b>	<b>113</b>
7.1	The Analyses . . . . .	113
7.1.1	Analysis I through III . . . . .	113
7.1.2	Analysis IV . . . . .	114
7.1.3	Analysis V . . . . .	116
7.1.4	Analysis VI . . . . .	117
7.1.5	Analysis VII . . . . .	117
7.2	Observations . . . . .	118
7.2.1	The Research Questions . . . . .	119
7.2.2	Reference Sample Sizes . . . . .	120
7.2.3	Selecting Landmarks . . . . .	120

7.2.4	The Importance of Morphological Distance . . . . .	120
7.3	Comparison to other results . . . . .	122
7.4	Problems . . . . .	123
7.5	Recommendations . . . . .	125
7.5.1	On choosing a reference sample . . . . .	125
7.5.2	On choosing the technique . . . . .	125
<b>8</b>	<b>Conclusion</b>	<b>129</b>
8.1	Summary of the Aims . . . . .	129
8.2	Overview of the Testing and Results . . . . .	130
8.2.1	Exploratory Testing of the Reconstruction Methods: Analysis <i>I–III</i> . . . . .	130
8.2.2	When Does Regression Become Viable: Analysis <i>IV</i> . . . . .	130
8.2.3	How Does the Amount of Damage Effect Residuals: Analysis <i>V</i> . . . . .	131
8.2.4	Do Landmarks Benefit Equally From Reconstruction: Analysis <i>VI</i> . . . . .	131
8.2.5	How Well Do Fossil Specimens Fare Under Reconstruction: Analysis <i>VII</i> .	131
8.3	Contributions . . . . .	132
8.4	Future Work . . . . .	133
<b>A</b>	<b>Various Tables</b>	<b>137</b>
A.1	Tables For Analyses <i>I</i> through <i>III</i> . . . . .	137
A.2	Tables For Analysis <i>IV</i> : Cross Over . . . . .	141
A.3	Tables For Analysis <i>V</i> : Cumulative Errors . . . . .	152
<b>Bibliography</b>		<b>163</b>

# List of Tables

1	Taxonomy of selected hominins . . . . .	3
2	Recorded landmarks . . . . .	90
3	Grouped Corrections of Fossil Specimens . . . . .	109
4	Corrections of fossil specimens . . . . .	110
5	Corrections fossil specimens, continued . . . . .	111
6	Corrections of fossil material using STS 5 . . . . .	112
7	Corrections of fossil material using KNM-ER 406 . . . . .	112
8	Corrections of fossil material using KNM-ER 3733 . . . . .	112
9	A comparison of results . . . . .	123
10	Equal reference size <i>t</i> -test results, mean substitution vs regression . . . . .	137
11	Unequal reference size <i>t</i> -test results, mean substitution vs regression . . . . .	138
12	Equal reference size <i>t</i> -test results, TPS vs regression . . . . .	138
13	Unequal reference size <i>t</i> -test results, TPS vs regression . . . . .	139
14	Equal reference size <i>t</i> -test results, TPS vs mean substitution . . . . .	139
15	Unequal reference size <i>t</i> -test results, TPS vs mean substitution . . . . .	140
16	Mean substitution and regression-based techniques as reference sample sizes increase	141
16	Continued from previous page. . . . .	142
17	Thin plate spline and regression-based techniques as reference sample sizes increase	143
17	Continued from previous page. . . . .	144
18	Thin plate spline and mean substitution techniques as reference sample sizes increase	144
18	Continued from previous page. . . . .	145
18	Continued from previous page. . . . .	146

19	<i>t</i> -test for small sample cross over: chimp reference . . . . .	147
20	<i>t</i> -test for small sample cross over: gorilla reference . . . . .	148
21	<i>t</i> -test for small sample cross over: human reference . . . . .	149
22	Mean residuals for figure 18 . . . . .	150
23	<i>t</i> -tests for cumulative errors . . . . .	150
24	<i>t</i> -tests for cumulative errors, regressions . . . . .	151
25	Mean residuals for cumulative error analysis . . . . .	153
26	Mean residuals for cumulative error analysis with 1 reference . . . . .	154
27	<i>t</i> -test for cumulative errors: mean substitution vs tps . . . . .	155
28	<i>t</i> -test for cumulative errors: mean substitution vs regression . . . . .	156
29	<i>t</i> -test for cumulative errors: thin plate splines vs regression . . . . .	157
30	<i>t</i> -test for small reference sample cumulative errors: chimp reference . . . . .	158
31	<i>t</i> -test for small reference sample cumulative errors: gorilla reference . . . . .	159
32	<i>t</i> -test for small reference sample cumulative errors: human reference . . . . .	160
33	<i>t</i> -test between cumulative errors: mean substitution . . . . .	161
34	<i>t</i> -test between cumulative errors: thin plate splines . . . . .	162

# List of Figures

1	A map of South Africa . . . . .	15
2	Typical formation of limestone caves. . . . .	16
3	Homology demonstrated with endocanthion. . . . .	20
4	The impracticality of affine transformations. . . . .	25
5	Offset of the optical axis. . . . .	39
6	Camera geometry . . . . .	40
7	Images from steps in the photogrammetric method. . . . .	42
8	Epipolar geometry . . . . .	44
9	A mid-sagittal CT slice through STS 5. . . . .	51
10	A render of STS 5 acquired from volumetric (CT) data. . . . .	52
11	Simplicies in 2D and 3D. . . . .	54
12	The Voronoi diagram and its dual, the Delauney triangulation. . . . .	55
13	Surface reconstruction of STS 5. . . . .	56
14	The medial axis. . . . .	57
15	MDS Rotation Independence . . . . .	65
16	The function $U(x, y) = -(x^2 + y^2) \log(x^2 + y^2)$ . . . . .	78
17	Various results for the techniques. . . . .	92
18	Comparison with increasing reference sample sizes . . . . .	96
19	Small reference sample comparison of techniques . . . . .	97
20	Cumulative error growth in the correction techniques . . . . .	99
21	Cumulative error growth with small reference samples . . . . .	100
22	Error distributions for the mean substitution method. . . . .	105

23	Error distribution for the thin plate spline method . . . . .	106
24	Error distribution for the regression-based method . . . . .	107
25	A flow chart of recommendations . . . . .	127

# Chapter 1

## Introduction

Even with the advent of molecular techniques, human evolutionary history is still studied primarily through the interpretation of the fossil record. Unfortunately, fossils are not perfectly preserved remains of the organisms we wish to study: they are remains that have undergone post-mortem changes to reach us in a state very different from that which the organisms held in life. These changes effect how easily we can reason about and understand extinct species.

The most frequent and well known post-mortem change is the chemical transformation of organic material into rock — but alongside this occur many physical post-mortem changes which effect the fossil materials' actual shape. Unlike chemical transformation, these physical transformations are of great interest to the study of morphology.

This interest in physical transformation can be seen in the growing body of work concerning the reconstruction of fossil fragments and the removal of plastic distortion. This thesis attempts to strengthen research in the mathematical reconstruction of fossils by comparing the behaviour of three reconstructive techniques, two of which (*mean substitution* and *multiple linear regression*) are *statistical* in nature, with the third (*thin plate splines*) being *geometric*. These techniques are compared by using each method to reconstruct the same set of individuals, then examining the resulting error residuals; this is done while manipulating various independent variables to investigate their effect on the behaviour of these techniques. Two of the main thrusts in our empirical analyses concern how the size of the reference samples — the sample of individuals that make up the prior knowledge used to correct a damaged fossil — and the taxonomic affinity of the reference sample (relative to the damaged specimen) effects the outcome of the reconstruction.

An important question that this work attempts to answer is whether the morphology of living (*extant*) species can be used to accurately reconstruct fossils of extinct species. Because relatively few undamaged fossil specimens exist, the consideration of whether large reference samples drawn

from extant species can be used in lieu of extremely small samples of extinct species — and when it makes sense to do so — is important to fossil reconstructions.

The next few sections briefly introduce fossil hominins (our bipedal ancestors) and provide some context for this thesis. Section 1.3 describes and motivates this work, giving its aims and research questions. The results and contributions are outlined in 1.4. The chapter concludes with an outline of the rest of this dissertation.

## 1.1 The Fossil Record

The human evolutionary lineage began between 8 and 5 million years ago (mya) (Ruvolo, 1997), when human ancestors diverged from those of our closest living relative, the chimpanzee. Most of the evidence concerning the evolutionary history of our species, our direct ancestors and close relatives (an evolutionary group colloquially known as the *hominins*), is provided through the fossil record (Kidwell and Holland, 2002; Wood, 2000). The bulk of the known record spans the last 5 million years. Older, possibly hominin, fossils do exist, such as the *Sahelanthropus tchadensis* material (Brunet et al., 2002) from Chad, dated at 7 – 6 mya.

Regardless of when our earliest ancestors evolved, it is clear that they evolved in Africa, as all fossil evidence of hominins prior to approximately 1.8 mya comes from this continent. Indeed, most of the hominin remains were found in two areas: eastern Africa and South Africa.

Our best understanding of these early hominins has come from a relatively well-sampled subgroup known informally as *australopiths*. They differ from the species in our own genus, *Homo*, by generally having the body size of a chimpanzee, and a comparable, if not sometimes slightly larger, brain size (especially if *Homo habilis* is moved to the *Australopithecus* genus as has been suggested by Wood and Richmond (2000) and others). Unlike *Homo*, australopiths typically did not make use of tools (although *Australopithecus garhi* remains are associated with animal bones showing signs of being defleshed with tools (Asfaw et al., 1999)).

The exact number of australopith species is open to debate, as are the number of species in our own genus, *Homo*. Table 1 gives a brief overview of the hominin clade.

## 1.2 Taphonomic Distortion

As will be further discussed in the next chapter, many of these early fossil hominin remains suffer from post-depositional distortion. One well known example of this occurs in the South African context. The *Paranthropus robustus* and *Australopithecus africanus* material from the South African

		Hominoidea					
		Hominidae					
		Homininae					
Superfamily	Hylobatidae	Ponginae	Gorillinae	Panini	Hominini		
Family				Pan	Australopithecina	Hominina	<i>Homo</i>
Subfamily							
Tribe							
Subtribe							
Genus	<i>Hylobates</i>	<i>Pongo</i>	<i>Gorilla</i>	<i>Pan</i>			
Species		<i>P. pygmaeus</i> <i>P. abelii</i>	<i>G. gorilla</i> <i>G. beringei</i>	<i>P. troglodytes</i> <i>P. paniscus</i>	<i>Ardipithecus kadabba</i> <i>Ardipithecus ramidus</i> <i>Australopithecus afarensis</i> <i>Australopithecus africanus</i> <i>Australopithecus anamensis</i> <i>Australopithecus bahrelghazali</i> <i>Australopithecus garhi</i> <i>Kenyanthropus platyops</i> <i>Orrorin tugenensis</i> <i>Paranthropus aethiopicus</i> <i>Paranthropus boisei</i> <i>Paranthropus robustus</i> <i>Sahelanthropus tchadensis</i>	<i>H. sapiens</i> <i>H. neanderthalensis</i> <i>H. heidelbergensis</i> <i>H. ergaster</i> <i>H. erectus</i> <i>H. habilis</i>	

Table 1: One taxonomy of selected hominins (largely based on Wood and Richmond, 2000) showing the relationship between some of the extant primates as well as our close ancestors and relatives. The genus *Pan* contains the chimpanzee and bonobo. Strong evidence suggests that the chimpanzee is our closest living relative. Indeed, species in *Pan* appear to be more closely related to our own species, *H. sapiens*, than to either gorilla (*Gorilla*) or orangutan (*Pongo*). For this reason, *Pan* is included in our subfamily, Homininae, while *Pongo* and *Gorilla* are both excluded. The *Hylobates* genus consists of four subgenera, with species such as the siamang and various forms of gibbons, none of which have been listed here.

australopith fossil collection are generally found in limestone cave deposits in the Gauteng province of South Africa. While these caves have acted as areas in which predators could accumulate australopith remains (Brain, 1981), and have protected these remains from the elements, the caves have also been the main cause of much of the damage seen in the South African fossil material. In some cases, the sediments which have collected in their interior cover the fossils by as much as 30 m (Brain, 1981). The weight this places on the fossils has at times been greatly increased by erosion, which caused the subsidence and collapse of the cave roofs. This, combined with the fact that crania are essentially hollow spheres, has caused a range of plastic deformations in the South African fossil crania. The possible distortions can be extremely light, perhaps with only a mild deformation to certain features, while the overall shape and size — the cranium's *morphology* — remains relatively undamaged. However, the damage may be severe enough that any morphological study is restricted to examining only small, isolated areas not strongly effected by this *plastic distortion*.

This damage of fossil material falls within the field of *taphonomy*. Taphonomy is the study of the changes that an organism's remains undergo after death, including the different fossilisation processes that occur following burial. Understanding the possible taphonomic processes and their effects on a fossil specimen is an important step in correctly reconstructing both the original organism and its community. If taphonomic effects are not taken into consideration, it could lead not only to a restriction in our ability to study fossil material, but also to erroneous conclusions.

## 1.3 Motivation, Aims and Research Questions

### 1.3.1 Motivation

The rise of the study of taphonomy, driven by people such as C. K. Brain, has taught us caution when interpreting fossil specimens. It is important to know both if, and how, the material has been damaged. "If" is often not a difficult question to answer, at least for gross morphology: it is fairly easy to tell if the material has been fractured, and plastic distortion usually causes the specimen to violate the biological patterns that we expect primates to display, bilateral symmetry (i.e., the mirroring of the left and right halves of the body) being an obvious example.

To take these distortions into account in our studies — whether our studies be statistical or otherwise — is often a more difficult endeavour. This difficulty is increased by the lack of undamaged fossil specimens on which to base our reasoning. For instance, the fossil material making up the *Australopithecus africanus* collection (hypodigm) contains only one complete, undistorted adult cranium: STS 5 (Mrs. Ples). Still, many techniques for reconstructing damaged crania exist. These may be categorised as follows.

1. *Physical reconstruction* is the traditional method. Reconstruction is performed either directly on the fossil material, or on casts thereof. Fossil fragments are placed into their presumed correct anatomical positions, while multiple individuals are often used to fill in missing portions, leading to what are called *composite reconstructions*. The Le Moustier 1 cranium of a Neandertal child is one such reconstructed specimen: the cranium is fractured, displays plastic deformation, and has undergone numerous physical reconstructions since its discovery (Dieck, 1923; Klaatsch, 1909; Klaatsch and Hauser, 1908; Ponce De León and Zollikofer, 1999; Schuchardt, 1912; Weinert, 1925).
2. *Virtual reconstruction* has become popular with the introduction of CT devices and cheap computing power. It has a number of advantages over traditional methods: it can be performed without directly involving any specimens, thus helping to preserve fragile fossil material; it does not require support structures — such as struts or putty — to hold fragments in place; it is possible to artificially segment distorted portions of any specimen, and to place each portion into a more anatomically parsimonious position, thereby partially removing plastic distortion. Other advantages include the ability to scale and shear distorted material, but this is only partially successful in correcting deformation. Early virtual reconstruction methods can be summarised as representing essentially a shift of traditional techniques to a digital domain; authors took only limited advantage of the extra flexibility — in terms of mathematical and computational power — offered by the digital medium. The work of people like Glenn Conroy (e.g., Conroy et al., 1998, 2000) and the team of Christoph Zollikofer and Marcia Ponce de León (e.g., Ponce De León and Zollikofer, 1999; Zollikofer et al., 1995) exemplify this approach.
3. *Morphometric based reconstruction* is a more recent development in virtual reconstruction. Morphometrics is the statistical study of shape, and the computing resources available to those using virtual reconstruction make a mathematical and statistical approach to reconstructing whole crania a viable option. Morphometric based reconstruction allows for the estimation of missing and damaged material based on a sample of known, undamaged reference individuals. It is capable of treating fossil material as if it were made of putty instead of stone, with the material hence being free to change form in nonlinear ways. A few authors have used morphometric techniques to “mold”, or warp, hominin specimens into a correct form. These methods are either guided by a researcher’s anatomical knowledge, as in much of Zollikofer and Ponce de León’s work — for example Zollikofer et al. (2005) — or more rigorous quantitative approaches are employed — such as in Gunz (2005) — where a single reference specimen determines the correct shape of the distorted crania. Morphometric based reconstructions subsume the techniques described as “virtual reconstruction”.

There is a certain amount of subjectivity inherent in all of these techniques. In physical and virtual reconstructions the knowledge required to perform the correction comes from the researcher's understanding of anatomy and physiology. In some cases this is the knowledge of a related species' anatomy, rather than from the species of the individual being reconstructed. This is true for the *SA-helanthropus* cranium reconstructed by Zollikofer et al. (2005), it being the only known example of the species (and genus). In contrast, morphometric techniques often place "knowledge" in explicit models, such as an "average" form calculated over a group of undamaged individuals, or perhaps a set of regression coefficients. In the worst case, only a single specimen is used to construct this reference model. However, these quantitative techniques have the advantage that the "knowledge" used to perform the reconstruction is easily shared with other researchers, and, if the correction technique is outlined with enough detail, the whole reconstruction should be reproducible by others with a minimum of subjectivity.

Of equal concern is the difficulty involved with attempting to correct plastic distortion. This difficulty is rooted in three problems:

1. Determining the compressional forces that have caused the damage is not straightforward, and any method that attempts to do this has to solve a woefully underspecified problem (Ponce De León and Zollikofer, 1999; Zollikofer and Ponce de León, 2005).
2. Statistical reasoning about morphology is centred around surface points that can be reliably identified across different specimens. Even if one were to determine how these *landmarks* should be corrected, it may not be clear how the correction should be transferred to the surrounding morphology.
3. Plastic distortion produces a non-affine deformation.

In reference to the first problem, researchers have generally avoided determining the causal forces of the deformation when performing reconstructions. Instead, distortion is dealt with by shifting fragments around or segmenting the cranium into artificial fragments (e.g., Conroy et al., 2000; Zollikofer and Ponce de León, 2005; Zollikofer et al., 2002a). Alternatively, the deformed anatomy is treated as missing, and reconstructed using mirroring or morphometric techniques (e.g., Gunz, 2005).

To address the second problem, researchers have used interpolation methods, such as thin plate splines (section 5.2.2), to distribute landmark corrections to the interlandmark morphology (e.g., Zollikofer and Ponce de León, 2005; Zollikofer et al., 2002a). In the case of reconstructing completely missing morphology, thin plate splines are used to warp another individual's morphology to fill the missing areas (e.g., Ponce De León and Zollikofer, 1999).

At all levels of the reconstruction, the question does not concern the *physics* (in terms of the kinetics, dynamics, and the physical makeup of the fossil material and surrounding sediments) of the damage and the correction implied by this, but rather a geometrical or statistical problem. This is not inherently bad — indeed, it transforms the problem into a tractable one. But one must be sure that the techniques used to perform the reconstruction, and the circumstances in which they are employed, are appropriate to the task at hand. For instance, consider the corrective power associated with physical and virtual reconstructions. These are generally limited to affine transformations (i.e., *rigid body* transformations, with *scales* and *shears*). While affine transformations are adequate to place fossil fragments into their correct position relative to one another, they are unable to completely remove plastic distortion, this being a decidedly non-affine form of damage. While affine-transformations have their use in fossil reconstructions, their limitations must be noted and addressed. However, the limitations of other techniques, such as thin plate spline methods, are not as clear, and their actual ability in predicting missing landmarks has so far not been examined in comparison to other techniques.

While the techniques looked at here (i.e., mean substitution, thin plate splines, and multiple linear regression) are more frequently finding use in the literature, it still remains unclear as to how suitable they are for performing accurate reconstructions. Often techniques appear to be used because they are both standard and expected (thin plate splines), or because received wisdom suggests one do so (e.g., to avoid the use of mean substitution). A deeper examination of the these techniques, including when and how they are appropriate, is required.

While these techniques do capture — in models of various kinds — the morphological knowledge that they employ, one must consider the amount of knowledge captured. When reconstructing fossil material we are almost always dealing with small sample sizes. There is only one known *Sahelanthropus* cranium, and when correcting it we have no other members of its species from which we can obtain knowledge of its morphology. This problem is not unique to *Sahelanthropus*. Inevitably, researchers employ their more generalised knowledge of primate morphology, this essentially being a morphological knowledge of *other* species. Can this *across-species* knowledge be used in the more formal setting of morphometric reconstruction techniques? This question, along with the need for an examination of the behaviour of the correction techniques, leads us on to the aims of this thesis. These are discussed in the following section.

### 1.3.2 Aims and Research Questions

This thesis presents work comparing three mathematical reconstruction techniques that have previously appeared in the literature. It attempts to show how the techniques perform relative to each

other in an empirical manner, which will hopefully allow others to make a more reasoned judgement on which techniques to apply, and when.

The three techniques examined in this thesis are:

- *Mean Substitution*: Replaces a missing landmark with an average for that landmark, drawn from other undeformed specimens (section 5.2.1).
- *Thin Plate Splines*: Warps an undamaged individual onto what remains of the damaged specimen. Missing landmarks are then substituted with their warped counterparts (the *homologous* landmarks) from the undamaged individual (section 5.2.2).
- *Multiple Linear Regression*: Regression models relating landmarks to one another, calculated on undeformed specimens, are used to fill in a damaged individual's missing landmarks (section 5.2.3).

## Aims

The main aims of the thesis are:

1. to compare the corrective properties of *mean substitution*, *thin plate splines*, and *multiple linear regression*;
2. to examine how a lack of reference individuals effects the properties of the techniques in relation to each other;
3. to determine if it is wise to use extant species as a reference population, thus providing a large reference sample as opposed to the extremely small sample set from extinct species.

## Research Questions

A large portion of this work is exploratory. For instance, while we expect the regression-based method to outperform the others, its dependence on large reference sample sizes means that a part of the analyses in chapter 6 will be concerned with discovering just what sample size is required for the regression-based method to outperform both the mean substitution and thin plate spline methods. Nevertheless, some expected outcomes can be posed:

### 1. Amount of Error

*This concerns the average amount of expected error — relative to the other techniques — for correcting any one particular landmark.*

From previous work by Gunz (2005), we expect that mean substitution should perform worse than the other techniques. The thin plate spline method will perform better, while the regression-based method will outperform both of these.

Performance is measured by mean residuals (the mean of the difference between the true and corrected landmarks). Mean substitution should obtain a statistically significant larger mean residuals than the thin plate spline method, while, in turn, thin plate splines are expected to have a significantly larger mean residuals than the regression-based method.

## 2. Growth in Error

*This concerns how the error associated with a landmark correction grows as larger portions of the specimen require reconstruction.* We use the number of landmarks requiring simultaneous correction as a proxy for the amount of damage a specimen suffers.

The mean residual associated with mean substitution is not expected to significantly differ as the number of corrected landmarks increases, since mean substitution makes little use of non-damaged morphology. The thin plate spline and regression-based methods are, on the other hand, expected to show a statistically significant *increase* in residual sizes.

## 3. Small Reference Samples

*This concerns the estimation of landmarks when only a small reference sample is available.* This is, unfortunately, a typical case in palaeoanthropological work.

It is clear that regression-based methods are unusable with small reference samples. Mean substitution and thin plate spline methods can, however, be used. As before, the mean residual obtained with mean substitution is expected to be significantly larger than that obtained with thin plate splines.

## 4. Across-Species Correction

*This concerns what can be deduced about correcting a damaged individual using a reference sample drawn from a species other than that of the damaged individual.* Since the sample sizes of fossil material is relatively small, the variance / covariance patterns of living species are often used as a substitute for that of extinct species (as with Ackermann, 2002, 2005; Ackermann and Cheverud, 2000, 2002). In this vein, it is important to know whether reference samples from living species can be used to accurately correct members of extinct species.

Corrections performed with these *across-species* reference samples are expected to produce larger residuals than those performed with reference samples drawn from the same species.

- The first is concerned with obtaining detailed cranial models. The data used here was obtained using a mixture of photogrammetry (performed by the author), computed tomography (obtained elsewhere), and landmark acquisition using a contact digitiser (obtained elsewhere).
- Once the data was obtained, the question of how to process and display this had next to be answered. Various techniques are described, including the marching cubes (Lorensen and Cline, 1987) and power crust (Amenta and Bern, 1998; Amenta et al., 1998) algorithms for surface reconstruction, and point rendering algorithms (e.g., Rusinkiewicz and Levoy, 2000) for display.

## 1.5 An Overview of the Dissertation

This dissertation has two potential audiences: one from computer science and another from palaeoanthropology. This means that the dissertation has to be balanced by containing some background material that members of one audience will more than likely consider introductory. We ask that the reader bear this in mind.

The remaining chapters of the dissertation cover the following:

- Chapter 2 discusses the relevant taphonomic problems from fossilisation through to plastic distortion, fragmentation, and so on. Previous reconstructions are outlined, and problems with the techniques are discussed. Basic morphometric concepts are introduced.
- Chapter 3 presents various techniques for obtaining high resolution models. It briefly describes contact digitisers, laser scanning, and computed tomography. It then outlines the photogrammetric technique used to gather data for this thesis.
- Display and processing methods for the models are compared in chapter 4. This includes methods for *point data*, as obtained by laser scanners and contact digitisers, and *volumetric data*, such as obtained from computed tomography.
- Chapter 5 describes how mean substitution, thin plate splines, and multiple linear regression are used in reconstructions. It describes some of the more unusual mathematics behind these techniques in more detail, such as multidimensional scaling.
- Chapter 6 describes the methods and materials used in testing the techniques. The results are presented here and in appendix A.
- These results are then discussed in chapter 7, including how they relate to the research questions. Recommendations for researchers performing reconstructions of damaged fossil material are supplied.

- The thesis concludes in chapter 8. Here we discuss future work that could improve our understanding of the reconstruction techniques, and summarise all that has been previously discussed.

# Chapter 2

## Background

This chapter discusses fossilisation and taphonomy, describing South African cave taphonomy in detail. Eastern and central African sites are briefly discussed. The chapter also provides an overview of morphometrics. It covers previous techniques used in cranial reconstruction, especially in the difficult case of removing plastic distortion, then outlines various problems currently associated with the reconstruction of fossil hominin material.

### 2.1 Fossilisation

Organic remains are a fragile thing. The decay of soft material (muscle, brain matter, various internal organs and so forth) begins shortly after death, and if the remains contain no hardy substance — such as a vertebrate's mineralised skeleton — an organism is unlikely to leave behind any fossil trace at all.

Unfortunately, even for harder remains, the environment contains many factors easily capable of damaging and reducing their preservation (Lyman, 1994): predators and scavengers chew on and break exposed bone; plant roots and burrowing organisms damage and destroy buried bone. Animal trampling and exposure to the elements play their part. Bone itself eventually decays.

Because of these factors, the preservation of organic remains requires special circumstances to halt or delay their destruction. Invariably, the better preserved the remains appear over a given length of time, the more rare were the environmental conditions leading to this state (Kidwell and Holland, 2002; Lyman, 1994). Even if any remains survive into the present without suffering physical damage, the fossil material has often undergone chemical alteration through the leaching of chemicals from the bone, and their replacement with surrogates from the surrounding ground water. This is a slow process called *permineralisation*. The end result is the chemical transformation of the bone

into material similar to that of its environment; in other words, permineralisation transforms organic material into stone.

As a rule, post-mortem effects generally alter the information content of fossilised material. Bearing this in mind, many people have recognised the usefulness of understanding these effects and how they alter our ability to reason and draw conclusions about fossil remains. This is the domain of *taphonomy*.

## 2.2 Taphonomy: “The Laws of Burial”

There are two areas of concern when considering the changes that fossil material undergo (Lyman, 1994):

1. *Changes that accumulate between an organism’s death and its burial.* This includes, for example, the action of predators and scavengers and the disarticulation of the various bones as the soft tissue holding them together decay. These *pre-depositional* effects fall within *biostratinomy*. The question of why fossil remains have collected in a particular area are an important part of biostratinomy.
2. *Post-depositional changes.* Permineralisation is itself a post-depositional effect, as are the compressional distortions and other factors that this thesis is concerned with. Other examples include damage caused during collection and curatorship: for example, some of the South African fossil accumulations are associated with limestone mines. Blasting at the mines has damaged some material in various ways (Brain, 1981). Mrs. Ples (STS 5) shows such damage: the specimen’s calvaria has been neatly detached from the rest of the cranium (Brain, 1981). The study of post-depositional changes, especially chemical changes, makes up the field of *diagenesis*.

Together, diagenesis and biostratinomy make up the science of *taphonomy*: the study of how the fossil record relates to the original biotic communities whose dead<sup>1</sup> have reached us through various changes (Conroy, 1997; Lyman, 1994). This can only be done by taking into account the post-mortem changes that the fossil material has undergone.

While it is possible to correct for taphonomic effects through a mixture of sampling strategies and analytical techniques (Kidwell and Holland, 2002), there exists no “simple fix” for correcting

<sup>1</sup>We should probably point out to the reader that although this thesis is focused on the fossil remains of organic material, fossil remains need not always be so. The *Latoli footprints* are the tracks of three *Australopithecus afarensis* individuals, two adults and a child, left in volcanic tuff. These date from over 3 mya, and are a fine example of inorganic fossil remains.

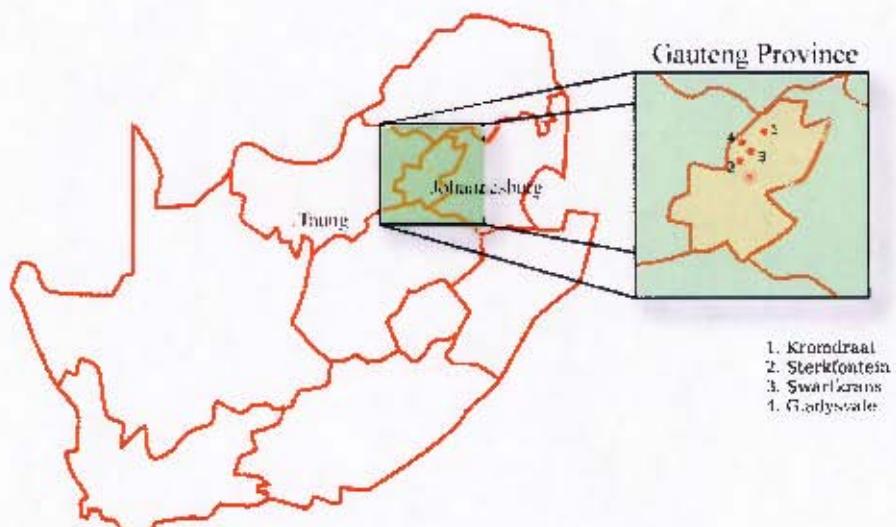


Figure 1: A map of South Africa, showing the sites that have supplied some of the earliest hominin fossil material.

all taphonomic effects; the solutions that a palaeoanthropologist, or a paleontologist in general, will require depend on the questions that they wish to answer given fossil material at hand.

### 2.3 Australopith Fossils in South Africa: Sites and Taphonomy

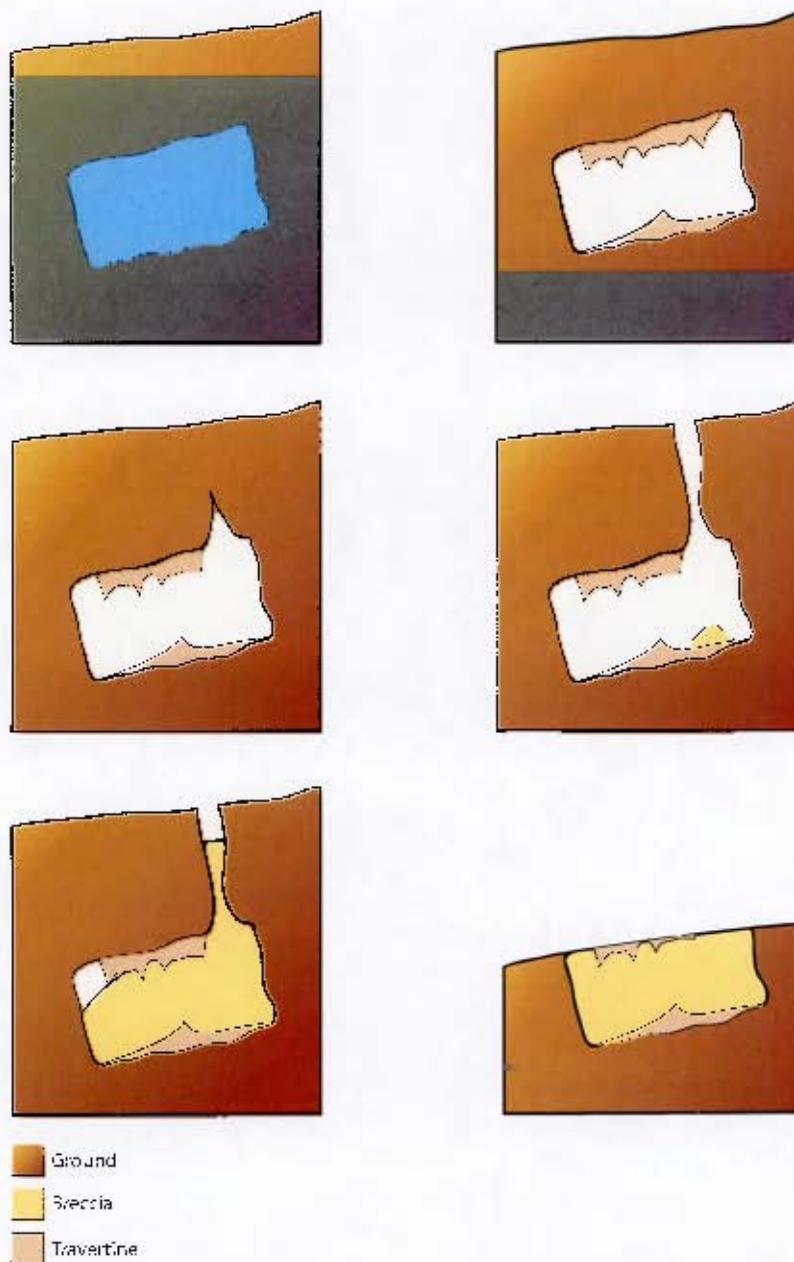
The first hominin discovery in Africa occurred in 1924 at the Taung limestone mineworks. All that remains of this is a partial *A. africanus* skull and an endocast<sup>2</sup>.

The *Taung child* material are the only hominin remains currently known from the site. However, various *A. africanus* material has been recovered from other South African localities, such as Sterkfontein, Makapansgat and Gladysvale. The species is currently known to have existed between 2.8 and 2.3 mya (Klein, 1999).

Material from the second South African australopith species, *P. robustus*, comes from sites such as Kromdraai, Swartkrans, Drimolen, and Gondolin. The species appears to have existed between 1.8 and 1 mya (Klein, 1999).

Without exception, all of these sites are associated with limestone caves. These caves generally appear to have been underground caverns connected to the surface by vertical shafts, with the main exception being Makapansgat, to which the entry appears to have been through a horizontal tunnel. Figure 2 on the following page demonstrates the typical formation of the limestone caves.

<sup>2</sup>An endocast is a cast of the internal morphology of the cranium. In this case, it was naturally occurring.



**Figure 2:** This sequence illustrates the typical formation of the limestone caves associated with the South African hominin fossil material (after Brain, 1981). The cave forms beneath the water table, but the table itself eventually recedes below the level of the cave, exposing the cavern to air. Travertine — deposits of calcium carbonate of which the more conspicuous examples are stalagmites and stalactites — is free to form. Any water flow through joints in the limestone erodes openings called *avens*. External matter falls through the *avens* to collect in the cave, forming sediments. Calcite bearing solution (the same that forms the travertine) transforms the sediments into cement like breccia. Erosion eventually exposes the sediments and breccia to the surface. This sequence does not include description of any taphonomic complications, such as subsidence or collapse of the cave roof, or the formation and collapse of secondary cavities.

Surface material (soil and such) enters and collects within the cave, forming sediments. These sediments may become cemented together by calcium-carbonate solution formed from the limestone; this creates *breccia*, an extremely hard “cemented” composite of heterogeneous materials: sand, silts, eroded portions of the cave, organic remains, and so on. Over its lifetime the cave both fills with sediments and suffers from erosion. This leaves — in the case of Swartkrans, for instance — fossiliferous breccias exposed to the surface (Klein, 1999).

Apart from exposing the breccias, erosion can cause both the subsidence and collapse of the cave roof. The cave floor can also collapse from the development of undercutting caverns, shifting the sediments laying in the upper cavern and resettling them. The compound result of these compressional forces is to distort the fossils contained within the sediments. This form of damage is widespread throughout the South African fossil material. For example, almost all the specimens from Member 4 at Sterkfontein have undergone some distortional effects of this nature (Brain, 1981).

The distortion of the crania is not necessarily uniform, such as a simple compressional effect restricted to only one dimension. This means that it may not be possible to describe the deformation using an affine transformation<sup>3</sup>. One important cause of these non-affine distortions is the heterogeneity of the fossiliferous sediments: larger bits of stone can create local indentations when forced into the cranium (Brain, 1981).

Further, the hollowness of the *cranial vault*<sup>4</sup> makes this area of the cranium more susceptible to distortion than the facial bones. However, the cranial vault also contains many openings, the chief entryway being the opening for the spinal cord, the *foramen magnum*. These openings allow the cranial vault to fill with sediments, increasing its resistance to the compressional forces of the overburden. However, it is not a given that the vault will completely fill with sediments. The orientation of the cranium, for instance, can cause the vault to only partially fill — the result of this is that filled portions of a cranium can withstand compressional stresses exceeding the rest of the vault, but the unfilled portions remain very prone to distortion (Brain, 1981).

Examples of compressional damage include the specimen SK 27 from Swartkrans. SK 27 is badly crushed (Clarke, 1977): the neurocranium itself has been flattened and the facial bones twisted. Other specimens have only been lightly damaged, such as SK 48, which had been almost completely filled with sediment, and so suffers only some damage to the occipital region of the brain case (Brain, 1981). SK 603 displays extreme shearing effects, caused by the shifting of the sediments containing the fossil after the collapse of the cave floor (Brain, 1981). Local compressional effects are found, for example, on specimens SK 79, which has a compressed nasal region.

<sup>3</sup>Or a series of affine transformations, since affine transformations form a multiplicative group: in other words, an affine transformation applied after another affine transformation can be described by a single such transformation.

<sup>4</sup>The portion of the cranium containing the brain.

Compressional deformation of this kind is sometimes avoided, although this does not happen often. Brain (1981) points out that STS 5 avoided all such effects, even though the cranial vault remained empty of sediment. The cranium came to rest close to the back wall of the cave, almost in contact with the roof. It managed to avoid the pressure of sediment buildup and roof collapse, even though fossils only a few meters away were damaged (Brain, 1981).

## 2.4 Australopith Fossils in eastern Africa

Apart from South Africa, the other “hotspot” for African hominin discoveries has been the great Rift Valley of eastern Africa. Several countries — Tanzania, Kenya, Ethiopia — offer rich deposits of fossil specimens, generally occurring in lake and stream deposits rather than cave deposits (Klein, 1999). Being associated with water, many of the fossils have been moved from their original resting position by water flow. However, there are localities which offer undisturbed fossil deposits from original living sites.

Many of the specimens have been discovered only after being exposed through erosion. Exposure to the elements have worn away portions of the fossil material, and the remaining material is often friable and fragmentary. When compared to the South African specimens, the east African specimens are often not distorted by compressional effects, while the South African specimens are generally not as fragmented.

One form of distortion that we find in the east African specimens that is not often seen in South African specimens is so called *expanding matrix distortion* (EMD) (White, 2003). This distortion is characterised by crania being highly fragmented; the cracks between the fragments are filled with varying amounts of matrix, with the matrix appearing to have pushed the fragments apart. The fragments are still held together by the matrix, so that they appear as one specimen. While plastic distortion tends to *crush* anatomy, Zollikofer and Ponce de León (2005) note that EMD tends to *expand* anatomy. However, both forms of damage are non-linear in their effect. The *Kenyanthropus* cranium is a well known cranium suffering from severe EMD; in fact, the distortion is so severe that even the cranium’s single tooth crown is effected (White, 2003).

## 2.5 Australopith Fossils in central Africa

Central Africa has so far not been a rich environment for fossil hominins, although considerably less palaeontological research has been carried out here. Still, the recently discovered Toumai (*Sahelanthropus tchadensis*) specimen from Chad suffers from plastic distortion (Brunet et al., 2002): “... almost the entire cranium has been flattened dorsoventrally [front to back] and the entire right

side is depressed.” We should note that, like much of the distorted material, the Toumai cranium also displays some undistorted portions. In this case, the damage appears not to have been related to cave formations, the specimen having been found embedded in sandstone showing aeolian (wind blown) and lacustrine (lake) deposition (Vignaud et al., 2002).

## 2.6 Morphometrics: Basic Concepts

The main quantitative approach to the study of an organism’s form and its relation to the world is to apply the tools of *morphometrics*. *Morphometrics* is a branch of multivariate statistics that focuses on how morphology covaries both with itself and with external factors (such as food intake, questions of biomechanics, etc.). Bookstein (1991) speaks of morphometrics being the study of a shape’s associations, its causes, and its effects. It is, essentially, the statistics of shape.

Traditionally, morphometric analysis makes use of only those variables best thought to represent the form in relation to the questions at hand. This information — usually only a few linear distances — is often information poor: although it can capture the information necessary for the study, the typical data set does not capture the complete “geometry” of the form (Richtsmeier et al., 1992). Such morphometric techniques are called *multivariate morphometrics*, to stress that the techniques are primarily only statistical in nature. Multivariate morphometrics transform the limited amount of collected data into various statistics, and while this data does have a graphical representation (plots combined with principal components analysis to reduce high dimensional multivariate spaces into a dimension amendable to plotting), the information contained in the statistics cannot be represented directly in terms of the original form (Zollikofer and Ponce de León, 2005). Various techniques have been developed that focus on obtaining *more* information from the individual’s form, followed by various analyses whose results can be easily transformed and represented in terms of this form (Zollikofer and Ponce de León, 2005). Indeed, Bookstein (1991) stresses that the method of choice for reporting results should be overlays on the original form data, rather than tables of statistical results. These techniques are able to more completely represent an individual’s morphology, both in terms of the data and the analytic results, and this — combined with the use of what is seen as “geometric” techniques — has led to the approach being called *geometric morphometrics*.

There are various geometric morphometric techniques, each with their respective utility. Many morphometric techniques — including so called multivariate morphometrics — are rooted in the study of *landmarks*. Landmarks are positions of anatomy that are easily identified across various specimens, and as such are usually named. These positions are the “same” position across individuals, in the sense that they are *homologous*: in other words, these are anatomical points shared due to common evolutionary descent. We say that two homologous landmarks form a *homology*.



Figure 3: The landmark on each face is *endocanthion*, the intersection closest to the nose between the upper and lower eye lids. These landmarks are *homologous*; they represent the “same” point on both people. The important concept behind homology and homologous landmarks is that they are the same due to evolutionary common descent.

For example, one of the many landmarks on the reader’s face is the intersection closest to the nose between the lower and upper eyelid (i.e., the *medial*<sup>5</sup> intersection between the upper and lower eyelid — see figure 3). This landmark is called *endocanthion*. Multivariate morphometrics deals with linear measurements between such landmarks, while geometric morphometrics use these landmarks in varied ways.

The collection of landmarks that are used to represent an individual are called the individual’s *form*. Thus an organism’s form is distinct from its morphology — the form representing only a discrete approximation to the true morphology. This thesis loosely speaks of correcting morphology — more accurately, what is meant is the estimation of missing landmark data from an organism’s form.

Geometric morphometric techniques include  
(Richtsmeier et al., 1992; Zollikofer and Ponce de León, 2005):

- *Finite-Element Scaling Analysis (FESA)*. A set of finite elements are constructed, these being subdivisions of the individual’s form into small, geometrical units using the landmarks as vertices. FESA measures the amount of strain produced on the elements by transforming one individual’s finite element model into that of another.
- *Procrustes Analysis and Kendall’s Shape Space*. This aligns landmark configurations from various individuals by performing a translation, rotation, and isotropic scale. The aligned landmarks are then averaged across individuals to construct a *consensus form*. Either the deviation between the original landmarks and the consensus landmarks are directly studied

<sup>5</sup>Closest to the midline.

via differencing (the traditional Procrustes analysis), or each individual's landmark configuration is merged into a single vector; the difference between the consensus and an individual's vector represents the deviation from the consensus form in a linearisation of *Kendall's shape space*. This space is a hypersphere of dimension equivalent to the merged landmark vectors. Analysis usually proceeds from a principal components analysis of the deviation vectors for all individuals under consideration.

- *Thin Plate Splines (TPS)*. Thin plate splines are interpolation functions, used to model the transformation of one set of landmarks onto another. This is similar to FESAs, with a notable exception being that it does not form a set of finite elements between the landmarks, and thin plate spline interpolations attempt to minimise a quantity called *bending energy*, which we discuss shortly.
- *Euclidean Distance Matrix Analysis (EDMA)*. The heart of this method is that landmark data is represented in such a way that it is invariant under rotation, translation and scaling, thereby avoiding the need for the alignment of data sets and its associated difficulties. EDMA represents a configuration of  $n$  landmarks implicitly as a  $n \times n$  matrix of all interlandmark distances, called the *form matrix*. Two forms are equivalent if their form matrices are equivalent (i.e., the distances between their landmarks are the same). Forms are the same up to scale (we say that their *shape* is equivalent, but not their *scale*) if their form matrices are proportional. The transformation of one form into another is represented by the *form difference matrix*, calculated by the component-wise division of one form matrix by the other.
- *Outline based methods*. There are anatomical areas that are landmark poor: they have few defining features that are identifiable across individuals. One way of analysing these areas is to study the form's outlines. This is done by a number of techniques, including transforming the outlines into frequency space using Fourier transformation. All analyses are then carried out in this frequency space. Semilandmarks (section 5.1.1) are a technique for making outline data amenable to landmark-based analyses.

These techniques are often combined. For instance, instead of studying form difference in shape space as a deviation from a consensus form, one can instead use thin plate splines to represent each form as a deformation from the consensus.

One can consider homologous landmarks to be a sampling of a "map" that relates one organism's form to that of another (e.g., Bookstein, 1991; O'Higgins, 2000). Landmarks are the only points of anatomy that are clearly identifiable across individuals: it would be difficult to locate the same arbitrary point on the neurocranium of two individuals if the point were not a landmark. Because landmarks are such easily identifiable "equivalences", the only sense that we can obtain of

this mapping is through these landmarks. Determining this *homology map* is thus an ill-posed problem: although the homology between landmarks defines how the map should act at the landmarks (homologous landmarks should be mapped to one another), what should occur to the interlandmark morphology is less clear. As such, Bookstein (1991) reminds us that we cannot speak of *the* homology map between two organisms, but can only speak of *a* homology map.

The use of *thin plate splines* is a common method of calculating a homology map (Bookstein, 1989, 1991). Given two forms  $\{x_i\} = X$  and  $\{y_i\} = Y$ ,  $X$  and  $Y$  being finite subsets of  $\mathbb{R}^3$  (our landmarks), a family of thin plate splines are used to define a map  $f : X \rightarrow Y$ .  $f$  displays a number of useful properties, perhaps the most important being (Bookstein, 1989, 1991):

1. Each  $x_i$  is mapped to its homologous  $y_i$ , ensuring that homologous points remain homologous.
2.  $f$  is the transformation between the two sets that has minimal *bending energy*. In other words, if the spline  $f$ , which takes as input the 2D point  $(x, y)$ , then  $f$  minimises

$$\iint_{\mathbb{R}^2} \left[ \left( \frac{\delta^2 f}{\delta x^2} \right)^2 + 2 \left( \frac{\delta^2 f}{\delta x \delta y} \right)^2 + \left( \frac{\delta^2 f}{\delta y^2} \right)^2 \right] \quad (1)$$

This is the integral of 2<sup>nd</sup> (i.e., curvature) derivatives. The minimisation of this “bending energy” is synonymous with how a thin sheet of metal minimises the amount of bending that it undergoes when subjected to deformation, and hence the name given to the splines. Section 5.2.2 describes thin plate spline warping in more detail, including the calculation of  $f$ .

## 2.7 Previous Reconstruction Examples

### 2.7.1 Overview

Fragmented crania have been corrected both directly using physical remains (i.e., on the fossil specimen, or their casts), and by means of computerised models. The main difference between physical and virtual reconstruction is that virtual reconstruction does not need to make direct use of the fossil material. It is capable of using data obtained through techniques such as CT (see chapter 3).

This section will not discuss physical reconstruction — it focuses, instead, on virtual and morphometric reconstruction. It begins by describing previous examples of such work, and ends with a discussion of some of the current problems.

method — are one way of capturing variance / covariance structure in our reconstructions. Doing so appears to reduce the overall landmark estimation error (see the results of Gunz, 2005). However, regression-based methods cannot be used with small reference samples. Indeed, they require samples large enough that they so far have not been widely employed in the literature.

### Problems with Error Estimation

To further this discussion, let us assume the existence of a certain test which we will call the *error estimation* test. Input to the test is (i) a known, undamaged, digital cranium; (ii) a list of landmarks to be *removed* from the cranium; and (iii) a correction technique. The correction technique is then used to predict the removed landmarks. The residuals between the predicted landmark positions and their known true positions are calculated. To obtain some statistical rigour, the test is repeated on a sample of test individuals. The effects of various circumstance — such as reference sample size, amount of damage, and so on — can be examined by treating these as independent variables to be manipulated, then comparing the obtained residual means for significant differences with that of a control group. This error estimation test provides a sense of a technique's *accuracy* (mean error) and *precision* (standard error of the mean); the introduction of manipulated variables shows us the behaviour of the correction methods under the restrictions and realities of real life application.

Many of the reconstruction techniques provide little quantitative sense of how they perform. The more recent work of Zollikofer and Ponce de León provide quantitative results after each correction, showing how well their reconstruction relates to the anatomy of various primates (e.g., Zollikofer et al., 2005). This is, however, not a description of how well the technique as a whole works, but rather an indication of whether a particular reconstruction has given us results in line with what we expect to see.

Zollikofer and Ponce de León themselves do make use of the “error estimation” test (see Zollikofer and Ponce de León, 2005; Zollikofer et al., 1998, for example) in order to validate whether a given reconstruction methodology, employed on a particular specimen, performs as expected. For instance, Zollikofer et al. (1998) artificially fragmented an *H. sapiens* cranium in order to simulate the fragmentation and loss of fossil material that the Neandertal specimen they wished to correct suffered from. The *H. sapiens* cranium was reconstructed similarly to the Neandertal cranium, and the comparisons between the reconstruction and the undamaged cranium suggests that the reconstruction error fell within the same range as “anatomical departures from bilateral symmetry”. This was used as an indication of the viability of the Neandertal cranium’s reconstruction. However, the test was only performed on a single individual, and so offers little statistical significance to the accuracy of the reconstruction.

Zollikofer and Ponce de León (2005) provide the following guidelines for verifying any particular reconstruction (rather than a reconstruction technique as a whole):

1. *Repeatability* concerns having the reconstruction repeated by *different* users, and having each user perform the reconstruction multiple times. This reduces inter-user and user-specific error.
2. *Robusticity* involves repeating the reconstruction using different anatomical assumptions. This tests whether varying the assumptions used to perform the reconstruction do not lead to markedly different results.
3. *Verification* has the user set up damage similar to the individual being reconstructed. The intended reconstruction is then performed, allowing one to see how the results compare to the original. This gives us an idea of the technique's accuracy under the given circumstances.

Assuming that the technique has some precision (i.e., small standard error), the *repeatability* requirement is useful for reducing user error (this is particularly important for reconstructions which rely on a researcher's "morphological eye"). If the technique has no precision (i.e., widely varying results), then the *repeatability* requirement should easily pick this up. Robusticity is intended to ensure that the user does not perform a reconstruction based on an *unsound* set of anatomical assumptions, or anatomical assumptions that have little or no bearing on the reconstruction; doing so may unduly vary the resulting reconstruction. Repeatability and robusticity have little to say concerning the *accuracy* of the technique; consider, the technique may produce results that are reproducible but inaccurate, always displaying a particular systematic error. Repeatability would not pick this up, and robusticity would only do so if the inaccuracy results from the anatomical assumptions rather than from a flaw in the technique as a whole. The *verification* guideline's goal is accuracy, and has a similar feel to our "error estimation". Still, we note that most published work does not supply much quantitative information in this regard.

Gunz (2005) provides some results concerning the accuracy of the various techniques he examines. Indeed, his thesis employs the "error estimation" test. The results from the test are interesting: the correction techniques under consideration performed increasingly poorly as the amount of damage on a cranium increased. They performed better when correcting simple forms — such as the smooth surface of the cranial vault — over complex forms — such as the facial bones. While Gunz (2005) provides a box plot of root mean squared residuals, the value of these RMS residuals are not supplied; neither are mean error<sup>11</sup> or the standard error of the mean.

<sup>11</sup>Note that root mean squared values are not the same as mean values, and exaggerate the effect of outliers (being the sum of squares).

While researchers have sometimes quantitatively considered the applicability of their reconstruction techniques, they often do little to show how they perform relative to other approaches. One way for the community to be sure of the techniques and their reconstructions — whether produced manually, using anatomical and physiological knowledge, or using mathematical and statistical procedures — is through the publication of quantitative results concerning mean errors and standard deviations of the observed mean errors with respect to key aspects such as the amount of area corrected, the type of area corrected, and so on. This should allow for more informed decision making when reconstructing fossil material.



## Chapter 3

# Data Acquisition

The data for the analyses carried out in later chapters was acquired from disparate sources; while one source was *computed tomography*, the bulk comes from photogrammetry. This chapter describes these sources and then goes into a detailed coverage of the photogrammetric method employed.

The chapter begins with an outline of the main data acquisition techniques available to the researcher. It then goes on to describe, in more detail, the photogrammetric technique that was employed to obtain some of the data used in this thesis.

### 3.1 Acquisition Techniques

A number of techniques have been used for imaging fossil specimens. *Radiography* (creating images by using x-rays) has been used by paleontologists for over a century (see Spoor et al. (2000) and Wood (2000) for examples), while the recent development of *computed tomography* (CT) has been an important advance in the detailed imaging of fossil specimens. Today, CT is the most widely used imaging modality for bony material in the palaeoanthropological community. Early use can be found in, for example, Conroy and Vannier (1987, 1989), or more recently in Conroy et al. (1998, 2000); Ponce De León and Zollikofer (1999); Zollikofer et al. (2002b).

However, CT and related radiological acquisition methods are not the only techniques available for acquiring models of fossil material. Methods worth considering include:

- Contact digitisers
- Diagnostic Radiology; specifically *computed tomography*.
- Laser scanning

- Photogrammetry

This chapter briefly discusses each method, along with obtainable resolution, and possible advantages and disadvantages.

### 3.1.1 Contact Digitisers

Contact digitisers provide the coordinate of any point that the digitiser is in contact with. There are two types of commonly used contact digitisers:

1. *Mechanical digitisers* measure the angles between various segments of a mechanical armature supporting the digitising pen. Using both the angles and the known segment lengths, the position of the pen tip can be calculated.
2. *Magnetic digitisers* measure the position of a magnetised tip within a larger magnetic field.

These instruments can achieve high degrees of accuracy, with, for example, some models in Immersion's *Microscribe*<sup>1</sup> range of digitisers having accuracies up to 0.23 mm.

The main advantages of contact digitisers are:

1. When compared to many of the other imaging modalities, especially CT, the cost of obtaining a contact digitiser is rather minimal.
2. Compared to other techniques, a contact digitiser is relatively easy to use, even though repeated measures should be made in order to minimise user error. Still, a user can rapidly acquire landmark data.
3. Their portability makes it easy to take the digitisers to the specimen, rather than *vice versa*.

Their main disadvantages:

1. Contact with fossil material can be damaging. For example, repeated use of callipers for measuring interlandmark distances wears down fossil material<sup>2</sup>. Contact digitisers are likely to have a less destructive effect, considering that they need but one contact point, do not place opposable pressure on any structures, and do not need to be wiggled into place (as done, for instance, when using callipers to find the shortest distance between two points).

<sup>1</sup><http://www.immersion.com>. As of November 2004.

<sup>2</sup>Remember that fossil specimens are measured repeatedly by different researchers over many years; friction wear easily accumulates over the course of decades.

2. Contact digitisers only allow for coordinate measurement on exposed surfaces of the specimen. Its internal morphology, and any structures still embedded within the containing matrix, cannot be measured.
3. The measurement of a large numbers of points can become laborious, especially when large data sets are required from multiple individuals, and repeated measures are taken to reduce random measurement error.

### 3.1.2 Diagnostic Radiology

Diagnostic radiology is the arm of medical science that concerns itself with medical imaging. The more common medical imaging techniques include *computed tomography* (CT), *magnetic resonance imaging* (MRI) and *ultrasound*.

When imaging fossil material, however, the preferred radiological modality is CT (Spoor et al., 2000). Other data sources are less useful for studying skeletal morphology, and are correspondingly less often used. For instance, MRI is best suited to imaging soft tissue, and bones cannot be usefully imaged. However, the technique is still useful for comparative and functional analysis with extant species.

A typical medical CT scanner rotates both an x-ray source and a detector around the specimen. The attenuation of the x-ray is measured in slices through the specimen, and these ‘slices’ make up separate *CT scans*. When stacked, these slices can be interpolated to create a 3D model of the imaged specimen. Medical CT scanners can achieve a resolution of between 0.3 and 0.5 mm within each slice (Spoor et al., 2000). The resolution obtained between the slices is usually poorer, with an obtainable resolution between 0.5 and 1.5 mm. However, microCT scanners exist that can provide much finer resolutions, even to between 1 and 200  $\mu\text{m}$ , and often the resolution between slices is identical to the within-slice resolution.

CT’s advantages are:

1. CT acquires *all* morphology up to that permitted by both the resolution of the scanner and any scanning artifacts. This detail cannot be obtained through the use of, say, a contact digitiser. This includes internal structures, and skeletal morphology hidden from the researcher in general.
2. CT is completely non-invasive: it is possible to scan a specimen while it is still contained within its sedimentary matrix. This means that there is typically no need to risk damage to the specimen by removing the enclosing matrix, or to damage the specimen to gain access to its internal morphology.

The disadvantages of CT include:

1. The cost of CT machinery is expensive. However, the researcher typically uses CT facilities on offer by a given institution, and the cost of doing this varies, and may even be negligible.
2. Typical medical CT scanners are not portable; the scanner cannot easily be taken to the fossil, although portable microCT scanners do exist.
3. When their relative densities are similar, it can be difficult to segment the fossil material from its enclosing matrix. Human intervention is typically needed to do so successfully (Zollikofer et al., 1998).

See Spoor et al. (2000) for an introduction to, and examples of, the paleoanthropological uses of CT and MRT methods. Zollikofer et al. (1998) provides further CT related examples.

### 3.1.3 Laser Range Scanning

Like the contact digitiser, laser scanning is also restricted to measurements of visible surfaces only. However, the advantage of the technique is that the laser scanner requires no contact with the fossil specimen.

Terrestrial Laser scanners (as opposed to those used in air-borne surveying) can be divided into three families (Ioannidis and Tsakiri, 2003; Pfeifer and Lichten, 2004; Schulz and Ingensand, 2004):

1. *Active triangulation scanners*. These operate by passing a stripe (drawn by laser) down the object's surface. A camera, typically a CCD, measures the stripe's displacement, and in so doing determines surface coordinates. Triangulation scanners are typically used for close range work, below two meters, with possible accuracies of thousandths of a millimetre (Schulz and Ingensand, 2004). For example, Minolta's VIVID 910<sup>3</sup> laser scanner has an effective range of between 0.6 and 2.5 m, and a reported accuracy of 0.008 mm.
2. *Time of flight (or pulse) scanners*. These scanners measure the time taken for the backscatter of a laser pulse (as opposed to a continuous laser signal) to return to the scanner. The time is used to determine the point's distance from the scanner. Time of flight scanners offer effective ranges from a few meters to several hundred. Indeed, some scanners can measure distances over 1 km (Schulz and Ingensand, 2004). The resolution is generally in the range of millimetres to centimetres (Pfeifer and Lichten, 2004). An example of this kind of scanner is the Mensi GS200, which has a recommended range of between 1 and 100 m, with a resolution of 3 mm at 100 m.

<sup>3</sup><http://www.minoltausa.com/vivid/products/vi910-en.asp>. Accessed November, 2004.

3. *Phase-based scanners.* These scanners emit a continuous laser signal towards the surface. The change in frequency in the laser's backscatter is used to determine the distance to the surface. Phase based systems tend to have a much shorter effective range than pulse based scanners, although the time needed to perform a scan is much shorter (Pfeifer and Lichten, 2004). Submillimetre resolutions are obtainable. An example is the *Imager 5006* from Zoller+Frölich, with a resolution of 7.6 mm at 50 m, and a range between 1 m and 79 m<sup>4</sup>.

Considering their effectiveness for close range work, triangulation scanners appear to be of the most use for acquiring models from fossil specimens.

Laser scanners have the following advantages:

1. They require far less work to acquire coordinate measurements than, say, a contact digitiser.
2. Along with requiring less work, they also acquire coordinates far quicker than attempting to do so with either CT, contact digitisers, or photogrammetry.
3. Laser scanners are portable: they can be taken to the actual specimen.
4. No contact is required with the fossil.

Their disadvantages:

1. Like contact digitisers, laser scanners are unable to image hidden and interior morphology. The surface of the fossil needs to be completely removed from its containing matrix.
2. Laser scanners are expensive to both purchase and rent.
3. Each of the different classes of laser scanner perform measurements in sweeps across the surface, and the user does not control the exact placement of these measurements. This makes it impractical to obtain specific landmark coordinates; the needed coordinate must be inferred from the data provided by the scanner, usually by interpolation between nearby measurements.

Laser scanning has rarely been used in paleoanthropological studies. Examples include the work of Wood et al. (1998) and Aiello et al. (1998), in which laser scans were made of the articular surfaces of associated post cranial remains. This data was used to determine if the articulating surfaces of an individual match more closely than a random pair taken from two individuals, one surface from each.

<sup>4</sup><http://www.zf-laser.com>. Accessed January 2007.

### 3.1.4 Photogrammetry

Photogrammetry is a set of techniques for measuring objects and terrestrial landforms based on 2D images (Kasser and Egels, 2002; Mikhail et al., 2001). This is done by making use of the projective relationship (a perspective transformation) that exists between the image, the camera used for imaging, and the original object (Mikhail et al., 2001).

The technique has mostly been used for topographic analysis, but as the cost and level of technical knowledge needed to perform photogrammetric analysis has reduced (mostly due to the advent of digital technology), photogrammetry has found more uses in other fields of enquiry (Kasser and Egels, 2002), with examples ranging from archaeology (e.g., Drap and Long, 2001), architecture (e.g.,Debevec et al., 1996), medicine (e.g., Pilgrim, 1992) and industrial measurement — civil engineering, mining (e.g., Smit, 1997), construction of vehicles and ships, and in metallurgy (Smit, 1997).

In the computer science field, 'photogrammetry' is perhaps best known for modelling architecture, such as in the work of Debevec et al. (1996, 1998), and for combining geometric knowledge with the reconstruction process — again, see the work of Debevec, but also, for example, Poulin et al. (1998). Computer vision has worked — often independently from the photogrammetric field — on obtaining 3D structure from images (see Faugeras (1993), or most other text books on computer vision). This independence can be seen in the avoidance in computer vision of *bundle adjustments* (described later) for calculating object coordinates and viewing parameters. In contrast, the photogrammetric field considers bundle adjustment to be a fast and accurate method, and it is the algorithm of choice for computing these parameters. See Triggs et al. (2000) for a discussion.

The main equipment needed to perform photogrammetry is a good quality camera, either digital or film. Other equipment (such as a film scanner) is sometimes employed, but is not strictly necessary.

The advantages:

1. Photogrammetry is a non-contact imaging technique.
2. It is extremely easy to take the needed equipment to the specimen. Photographing the fossil material is generally also easy, although some forethought is required in order to control the lighting conditions and to ensure that all portions of the specimen are covered by at least two photographs. The positions the photographs are taken from — the *camera stations* — should also be considered.
3. The photographs can be used to not only model the fossil, but to obtain textures for the model.

The disadvantages:

1. Just as with contact digitisers and laser scanners, photogrammetry can only image visible portions of the fossil specimen.
2. Like the segmentation process for fossil material in CT, photogrammetry is not a completely automated process; someone is needed to perform the reconstruction process.
3. Semiautomated methods in photogrammetry can produce more noise than is associated with other techniques.

Photogrammetry was chosen to obtain the data used in this work. The reason for this was purely practical: the cost of performing CT scanning is considerable. Photogrammetry was considered to be less labour intensive than the use of a contact digitiser in terms of capturing interlandmark morphology, and a laser scanner was not easily obtainable. Software and expertise for performing photogrammetric analysis was also readily available to the author, and so these resources were utilised.

The following sections outline the photogrammetric analysis, and some accuracy results.

### 3.1.5 Data Forms

Each of the techniques described above provide 3D data. Each, however, provides them in slightly different forms. Contact digitisers, laser range scanning and photogrammetry provide only 3D points — this is essentially a *point cloud*, and further techniques are needed to create surface models from the data (see section 4.2.2 on page 53).

CT, on the other hand, provides *volumetric* data. Each scan is an image containing density information of a slice of the object. These slices are stacked, and isosurfaces — surfaces of the same density — contained in the stack can either be rendered directly, or extracted in some other form (for example, as a mesh via the marching cubes algorithm of Lorensen and Cline, 1987).

Some techniques can also be profitably combined. For example: laser scanning offers quicker data acquisition times than photogrammetry. Photogrammetry, however, offers good texture models. Photogrammetry can be used to determine the camera position for each photograph in relation to the scanned objects. This information can be used to “drape” the photograph over the laser scan model, thus producing both a detailed and textured model (see Bernardini and Rushmeier, 2002, for reviews).

The various data forms that the different input modalities supply are described further in section 4.1 on page 47.

### 3.2 Photogrammetric Outline

A photogrammetric analysis proceeds as follows: multiple images of an object are taken, with the images sharing overlapping areas of the object. This stereo overlap is central to photogrammetry: a single photograph contains no depth information (depth being relative to the camera); an object could lie anywhere on a line extending from the camera through the object. Using multiple photographs, each taken from different viewpoints (*camera stations*), the object's position can be triangulated.

The technique used here is that of Smit (1997). His method was developed for use in deep-level gold mines in South Africa. To lessen the stress of the rock mass being mined (these mines are located 3000 m below ground level), the area is pre-fractured through blasting. The technique was developed to allow modelling of the rock face before and after the blast, enabling the deformation that the rock face undergoes to be monitored. The advantage of this method is the use of automated interest point selection (i.e., the point whose 3D position is of interest) and the automated detection of their corresponding points in other images. In many other methods, corresponding points are identified by hand (for example, in Australis<sup>5</sup>, and Photomodeler<sup>6</sup>).

The photogrammetric reconstruction of crania can be outlined as follows:

1. *Establishment of a coordinate system.* All points lie within some coordinate system. In order to place object coordinates correctly within this coordinate system, the mapping between the camera's imaging geometry (the lens and imaging plane) and the real world object space needs to be determined.
  - (a) *Construction of a control frame.* This can be thought of as a portable coordinate system. It is a metal frame marked with points whose position are well known. This is used to determine the camera's position in each photograph.
  - (b) *Camera calibration.* The actual determination of the camera's imaging geometry characteristics is called *camera calibration*.
2. *Photographing the crania.* A series of photographs are taken in order to determine the surface coordinates of any given point. At the very least, two images are needed of every portion of the cranium.
3. *Calculation of the camera's external orientation.* This entails initially locating the centres of the control frame targets. Since the position of the targets are known, they can be used to calculate the camera's position in each photograph, up to and including scale.

<sup>5</sup>Commercial photogrammetric software from the University of Melbourne, Australia.

<sup>6</sup>[www.photomodeler.com](http://www.photomodeler.com) (as of December 2004). This is commercial photogrammetric package.

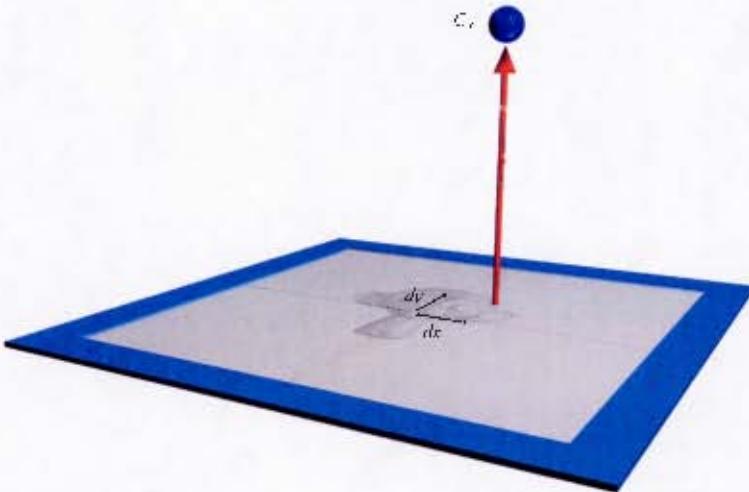


Figure 5: The optical axis (red line) does not necessarily pass through the image's centre, mostly due to lens aberrations.  $C_1$  is the camera's perspective centre.

4. *Locating interest points in a single image.* This is done automatically, using interest point operators.
5. *Locating the corresponding points in the remaining images.* Again, this is an automated process, using image cross-correlation.
6. *Calculating the object space coordinates of the interest points via intersection.* This entails the use of the *colinearity equations*.

Each step is covered in more detail below.

### 3.2.1 The Coordinate System

Photogrammetry requires the position and orientation of the camera to be known for each image, and this needs to be relative to some coordinate system. Smit's technique constructs this coordinate system before any image is taken. This position-orientation parameter pair is known as the camera's *external orientation*, and a physical *reference frame* is placed in each image to facilitate the calculation of these parameters. The reference frame (see figure 7(b) on page 42) is a physical object built from steel tubing marked with retro-reflective targets, with well known positions. The targets are 8 mm in diameter and show up well against the black paint of the steel frame. Since their coordinates are known, each target acts as a reference point for a coordinate system in every image that the frame occurs.

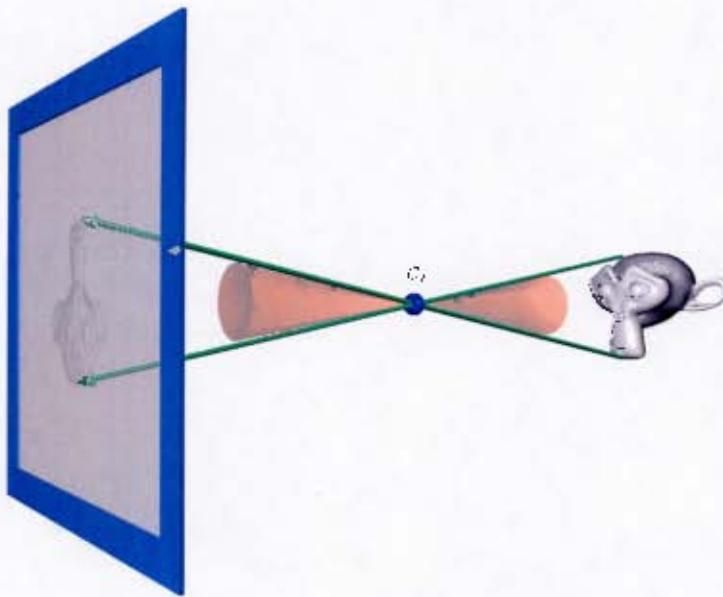


Figure 6: All rays that fall on the camera’s image plane pass through the camera’s perspective centre ( $C_1$ ). This causes the characteristic “flipping” of the image. The closest distance from the image plane to this point is the camera’s *principal distance*.

The reference frame target coordinates are determined using the same process as the *camera calibration*, described shortly. Apart from having a reference frame to determine the camera position, certain properties of the camera itself need to be known. *Camera calibration* is the process by which the geometric properties of the camera are discovered. These parameters include the lens’s *principal distance* (the distance from the perspective centre, the point through which the perspective projection occurs, to the image plane — see figure 6), and *principal coordinates* (the coordinates of the intersection between the camera’s *optical axis* — the line through the perspective centre and perpendicular to the image plane, shown in figure 5 — and the image plane). Cameras themselves are only approximations of an ideal pinhole camera, and so the imaging geometry does not perform a true perspective projection. For instance, because the lens takes up an area instead of being an infinitesimal point (Atkinson, 1996), field of focus effects occur during photogrammetric analyses that should be dealt with. Consequently an important part of camera calibration is the determination of how the camera geometry deviates from performing a true perspective projection.

These measurements are all performed by photographing a laboratory control field: a set of control points whose positions are very accurately known. Figure 7(a) on page 42 shows such a field. It is essentially a set of retro-reflective markers placed on a wall. The camera’s geometric parameters are solved by calculating provisional parameters using the known position of these targets through the use of a least squares method such as the *direct linear transform* (Kraus, 1997), followed by an optimisation of these parameters through the use of a *bundle adjustment*. A bundle adjustment is

the simultaneous least squares optimisation of the object coordinates with both the camera's internal and external parameters (Triggs et al., 2000). Essentially, the technique deals with the pencil of lines passing from the world being imaged through the camera's perspective centre. This pencil of lines can be thought of as a 'bundle', while the simultaneous optimisation of all the parameters can, also, be thought of as having everything "adjusted together 'in one bundle'." (Triggs et al., 2000).

Once these *internal orientation* parameters are known, the coordinates of the reference frame can be calculated by photographing it against the control field. The control field — whose point coordinates are known — can be used to determine the camera position. The camera station can again be calculated through the combined use of a direct linear transform and bundle adjustment, or through other methods such as that of Quan and Lan (1999). The reference frame target coordinates can then be triangulated from multiple images. The use of the control field is necessary, since the photogrammetric reconstruction cannot reconstruct the scale without some *a priori* information (such as reference points) (Debevec et al., 1996; Poulin et al., 1998). Because the position of the points in the control field is known, the scale factor can be calculated.

With the creation of the reference frame, and the calculation of the camera's internal parameters, the camera's position in each photograph can now be determined, as long as the reference frame is itself present in each image.

The commercial software package, *Australis*, was used to determine the calibration parameters.

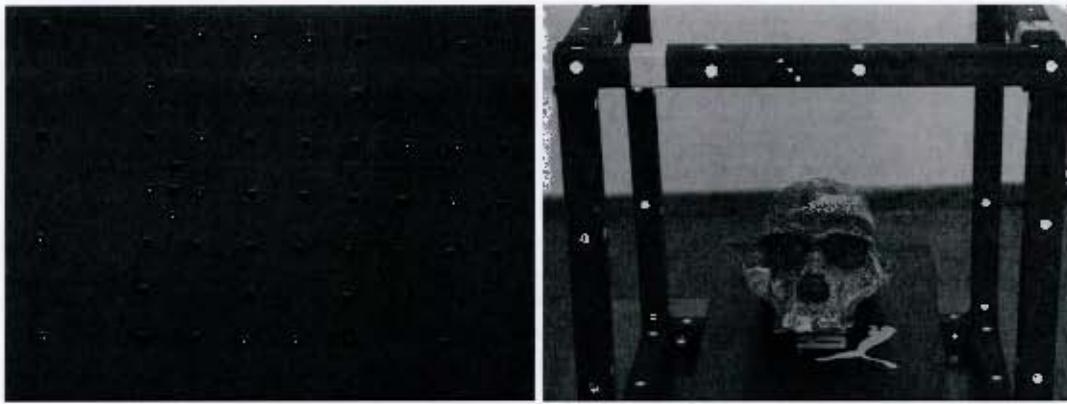
### 3.2.2 Photography

Each cranium was placed within the reference frame when photographed. Because of the reference frame's cube shape it was natural to photograph the cranium in six sets, one for each side of the cube. This creates six separate surfaces that are stitched together to create the final model.

Since the cranium remained in the same position relative to the reference frame while photographing each side, each surface remains in the correct position relative to the others.

The camera used for image acquisition was the Kodak DCS330, with the following features:

- SLR operation.
- 18.1 mm x 13.5 mm sensor area.
- 2008(h) x 1504(v) pixels.
- 2 x PCMCIA Type-II removable drive.
- Built in flash.
- 28mm lens.



(a) The calibration field. Multiple images of this field were taken from different camera poses, including variations in the camera's roll.

(b) The reference frame placed around a cast of STS 5. The frame is covered by retro-reflective targets allowing them to be easily seen by flash light. Multiple images, similar in nature to this one, were taken of each crania undergoing photogrammetric reconstruction.

Figure 7: Images from steps in the photogrammetric method.

### 3.2.3 Determination of External Orientation

Once the image acquisition is complete, the external orientation of the camera in each image needs to be determined. As previously mentioned, this makes use of the reference frame that appears in each photograph. As with the internal orientation parameters, *Australis* was used to calculate these parameters. The reader may wish to consult Quan and Lan (1999) for an effective camera pose estimation algorithm from the Computer Science literature.

### 3.2.4 Locating Interest Points

Smit's technique uses automated discovery of interest points for which 3D coordinate measurements are required. This is a considerable advantage over, say, the method of *Australis* or *Photomodeler*. These applications allow for determining the 3D position of a point, but *each* interest point and their corresponding positions in each of the images needs to be marked manually. As the number of points increase, manual point determination quickly becomes impractical.

The automated discovery of interest points is performed by running a *sobel interest operator* over one of the images. The Sobel operator marks interest points by detecting changes in an image's intensity gradient (Sonka et al., 1999). This does mean that the object of interest needs to have a surface with a varying intensity function — in other words, that it has a well textured surface. The texture variability of crania means that this requirement is easily met.

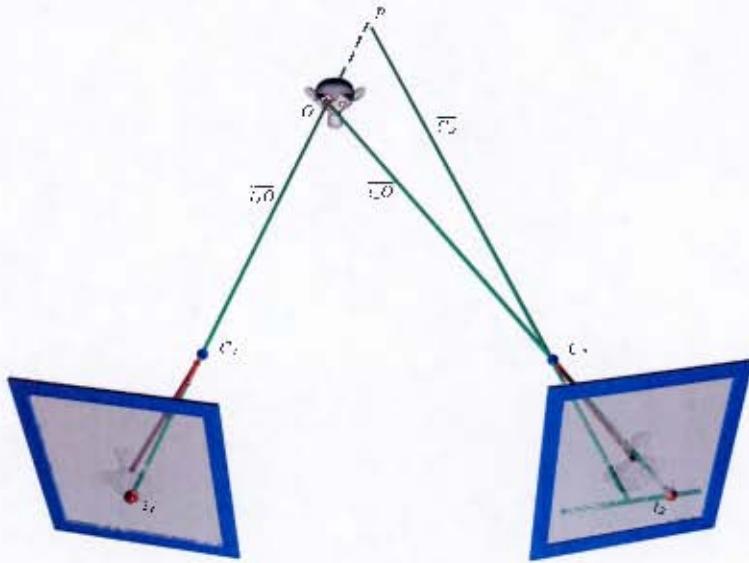


Figure 8: The epipolar line (the line on the image plane intersecting  $i_2$ ) is a constraint of object point  $O$ 's projection onto the image plane (this is  $i_2$ ), if the position of  $O$  is already known in another image (and this is  $i_1$ ). The projection of  $O$  must lie on this line. The red arrows are the camera's optical axes, while  $C_1$  and  $C_2$  are their perspective centres.

Once interest points have been marked in one image, their corresponding positions need to be located in the other images. These are found in an automated way. First, it is important to constrain the search space for this point. There are a number of standard techniques (Faugeras, 1993) of which Smit makes use (Smit, 1997):

1. *Epipolar Geometry.* This is perhaps the most important constraint, as it allows the search to be restricted to a one dimensional line across the image rather than a full 2D search throughout the whole image. Epipolar geometry is described below.
2. *Geometrical knowledge of the object.* For example, it can be assumed that the depth value of the object being measured does not rapidly change, and so larger step values can be taken in the search through the image.
3. *Physical constraints.* Often a Lambertian lighting model<sup>7</sup> is assumed across the surface of the object. This is because lighting (especially specular highlights) makes the surface texture of an object highly dependant on the viewing angle. Through the use of controlled lighting conditions, especially indirect rather than direct lighting, Lambertian illumination can often be approximated.

<sup>7</sup>This is a lighting model that takes into account only diffuse (i.e., *Lambertian*) reflections. It contains no specular highlights, or other such reflections.

### 3.2.5 Epipolar Geometry

*Epipolar geometry* describes the relationship between a point in one image and a corresponding line in another image. Consider two cameras each imaging an object (see figure 8). Of interest is the object point  $O$  which lies at point  $i_1$  on the image plane of the first camera, and point  $i_2$  on the image plane of the second camera.  $C_1$  and  $C_2$  are perspective centres for cameras one and two, respectively. The line between points  $i_1$  and  $C_1$  is the line along which the object point  $O$  lies. Call this line  $\overline{i_1O}$ . Note that it passes through  $C_1$  and is infinitely long (since depth cannot be determined by using only one image).

Consider camera two and its image plane. On constructing a line from  $i_2$  to  $O$  — call this line  $\overline{i_2O}$  — one can see that  $\overline{i_1O}$  intersects  $\overline{i_2O}$ , and that this intersection occurs at  $O$ . Notice that  $\overline{i_2O}$  passes through  $C_2$ , and extends to infinity.

If the projection of  $O$  onto the image plane of camera two is unknown (i.e., the value of  $i_2$  is unknown), one can proceed by constructing a line  $\overline{C_2}$  from  $C_2$  to an arbitrary intersection point  $p$  on line  $\overline{i_1O}$ .  $\overline{C_2}$  now intersects the image plane of camera two. If  $p$  is shifted along  $\overline{i_1O}$  (which corresponds to changing the estimated depth value of  $O$  as seen from camera one), line  $\overline{C_2}$  marks out a line across the image plane. This line is the so called *epipolar line*, along which the real position of  $i_2$  lies (Faugeras, 1993).

This epipolar line is searched by selecting a  $z$ -depth value and interpolating  $\overline{i_1O}$  at that depth. The *collinearity equations* specify how the object space coordinates along the line are converted to image space coordinates in image two:

$$x = x_p - dx + c \frac{r_{11}(X - X_c) + r_{12}(Y - Y_c) + r_{13}(Z - Z_c)}{r_{31}(X - X_c) + r_{32}(Y - Y_c) + r_{33}(Z - Z_c)}$$

$$y = y_p - dy + c \frac{r_{21}(X - X_c) + r_{22}(Y - Y_c) + r_{23}(Z - Z_c)}{r_{31}(X - X_c) + r_{32}(Y - Y_c) + r_{33}(Z - Z_c)}$$

$x$  and  $y$  are the image coordinates.  $x_p$  and  $y_p$  are the principal point coordinates,  $dx$  and  $dy$  are distortion parameters,  $c$  is the principal distance,  $X, Y$  and  $Z$  are the object space coordinates,  $X_c, Y_c$  and  $Z_c$  are the perspective centre coordinates, and  $(r_{ij})$  is the  $3 \times 3$  rotation matrix describing the camera's orientation. These equations give supply a candidate  $i_2$ .

The search continues from this candidate coordinate: Smit compares a patch around candidate  $i_2$  to a patch around  $i_1$  using the cross correlation function

$$R_{XY} = \frac{\sum(g_1 - \bar{g}_1)(g_2 - \bar{g}_2)}{\sqrt{\sum(g_1 - \bar{g}_1)^2} \sqrt{\sum(g_2 - \bar{g}_2)^2}}$$

The summation occurs over the pixels of each patch, and  $g_1$  and  $g_2$  are the grey scale values on images one and two, respectively, while  $\bar{g}_1$  and  $\bar{g}_2$  are, respectively, their averages.

The patches themselves are shaped to take into account the different positions and orientations of the two cameras.

The cross-correlation is evaluated at all the candidate points along the epipolar line, and the point whose cross-correlation values is closest to unity is chosen as  $i_2$ .

In this manner both the corresponding image point and the object space coordinate for the point are obtained. The process is repeated for each point that object space coordinates are required for.

This technique was used to obtain point data from the various hominin crania used in this thesis. Figure 13 on page 56 shows a surface reconstruction of STS 5 from data obtained using this technique.



## Chapter 4

# Data Processing and Display

Between acquisition and later analysis, morphometric data generally has to be processed to improve its usefulness. An important quality of the data used in morphometric analyses is that it represents real world shapes. The implication of this is that processing of the data should make it useful for display to the user<sup>1</sup>.

This chapter gives a brief overview of some of these techniques. The area is broad: we can only lightly touch the topics. They are, however, not central to the contributions of the thesis. The reader can easily follow up on any topics of interest through the supplied references.

### 4.1 The Form of the Data

There are two data forms that the various acquisition modalities discussed in the previous chapter supply. These are:

- *Surface data.* Intuitively, one can think of surface data as representing only what one can see of the object. It contains no information concerning the composition of these surfaces: surface data cannot tell us if the material is soft tissue or bone, whether it is fossil material or matrix infill. Measuring technologies that supply only surface data are usually unable to see beneath any given surface: they can measure only the visible superficial / external morphology, while hidden and internal morphology remains unrecorded. As discussed in chapter 3, most measuring techniques such as contact digitisers (section 3.1.1), laser range scanning (section 3.1.3) and photogrammetry (section 3.1.4) supply only surface data. There are two main representations of surface data:

---

<sup>1</sup>There is processing of a statistical nature that the data often undergoes, such as Procrustes averaging; such processing is dealt with in chapter 5.

1. *Point clouds.* This is without a doubt the most common form of acquired surface data. Contact digitizers, laser scanners and photogrammetry all provide point cloud data. Each point represents a single position measurement, and the point cloud lacks any information concerning the space between these measurements.
2. *Surface representations.* Point cloud data is usually post-processed to reconstruct an approximation of the surface from which the points were originally measured. This surface is described by a *surface representation* (see section 4.2.2 on page 53). A *surface representation* can be thought of as a  $d - 1$  dimensional structures lying in the  $d$  dimensional space that our measurements are from: planes in a 3D Euclidean space are a typical surface representation. A surface representation is often created by processing points to produce a collection of *planar polygons*<sup>2</sup> connected to one another at their edges. Planes can be thought of as *first order* — or *linear* — approximations to the surface, since the gradient vector of a plane<sup>3</sup> is constant. The reason for processing the data into a surface representation is to allow for data interpolation: point data cannot tell us about the surface between the individual point measurements, and so we interpolate between the points to approximate the missing information. Planar polygons used as a surface representation provides a linear interpolation between point measurements<sup>4</sup>. Different polygons making up the surface are free to have different gradient vectors.

*Higher order* representations allow the gradient vector to vary across any portion of the surface representation. These representations provide non-linear interpolation methods between the data points. The most common higher order representations are spline surfaces, such as B-spline and NURBS surfaces. Splines as a method of interpolation are discussed further in chapter 5, although we do not use them for representing surfaces. The interested reader can consult Farin (1992) for further information. The advantage of higher order representations is that the error between the original object and its surface representation is minimised without having to increase the number of polygons to better approximate the surface curvature.

Although most input techniques provide data initially as points, there are a few that avoid point clouds completely. As examples, consider the photogrammetric work ofDebevec et al. (1996), or Poulin et al. (1998). These techniques directly provide surface representations, but remain in the minority, with the most common measurement format for surface data still being the point cloud.

---

<sup>2</sup>Planes extending to infinity would not make for a useful surface representation. A planar polygon is, essentially, a bounded region of a plane.

<sup>3</sup>A plane has a gradient along both of its axes. Combining these gradients give a gradient vector rather than the simple scalar gradient that the reader is likely familiar with from calculus in one variable.

<sup>4</sup>Technically, it provides a *bi-linear* interpolation, since the interpolation occurs across two axes, rather than one.

- **Volumetric data.** Volumetric data supplies information concerning the composition of a region of space. Radiographs and CT scans are examples of volumetric data: both provide information concerning the density of the measured volume. A typical data form for volumetric data is to subdivide the sample space into numerous cells of equal dimensions, called *voxels*<sup>5</sup>. The voxels themselves have an *intensity* measurement associated with them. The intensity acquired through CT and other diagnostic radiology methods represents the density of the scanned object. Although it is possible to perform many operations directly on the voxel data itself, surface representations obtained from the voxel data are often operated on instead. Such a “surface” is calculated across user specified density levels (an *isovalue*), and represents a density *contour*, or *isosurface*. Modalities supplying volumetric data are able to measure the composition of an object, as well as image the internal morphology hidden to most surface acquisition devices. Isosurfaces, being defined by constant density, are useful for representing the boundary between, say, bone and soft tissue, or fossil material and matrix. It is important to realise that obtaining an isosurface from volumetric data only discards intensity information: the hidden surfaces belonging to internal morphology remain in the surface representation<sup>6</sup>.

## 4.2 Data Processing

While point cloud and volumetric data are inherently different, a surface constructed from both forms is still represented by the same underlying mathematics and data structures. Even so, the different data formats do require different surface reconstruction algorithms. Further, the interpretation of a surface itself varies with both the original data form and reconstruction algorithm. In the case of surface data, a surface representation is usually interpreted as an interpolation between data points. This can be complicated if the original data undergoes filtering (or an equivalent operation) before reconstruction to remove noise (*random error*). This tends to smooth the original data, and the reconstructed surface is then an *approximating* surface of the original data, rather than a true *interpolating* surface. On the other hand, a surface representation obtained from volumetric data is a display of an *isosurface*, or a surface constructed from areas of equal density<sup>7</sup>.

One problem to be mentioned before we proceed is the *sparsity* of the data, discussed by Amenta et al. (1998), Amenta and Bern (1998), and many others. To create an accurate surface,

---

<sup>5</sup>Although this regular grid is typical for data acquired from CT, volumetric data in its more general form need not be regular.

<sup>6</sup>As long as the density of the morphology lies in the range of the isosurface, otherwise it would not be included in the surface representation.

<sup>7</sup>Such surfaces themselves are also interpolations in a strict sense: they interpolate between the position of density measurements with the same or similar values.

both the volumetric data and the surface data must sample the specimen adequately. This is usually not a problem in terms of CT data, since the specimen is sampled fairly densely and uniformly (although the sampling rate may differ along different axes). Sparsity is of greater concern for surface data: data acquired through photogrammetry, for instance, is typically a random sampling from across the surface. This means that in some areas the point cloud may not be a dense enough sample of the object to accurately reconstruct the surface. This will be explained more fully in section 4.2.2.

Non sparse (*dense*) point clouds increase the “accuracy” — or the “fidelity” — of the surface reconstruction. This is due to the extra information concerning surface curvature and topology that a dense point cloud contains. Still, accuracy is closely tied to the reason for creating a reconstruction. If one only wants to study gross, superficial morphology, it is often a simple matter to sample the surface only as densely and uniformly as needed to create a surface capable of answering one’s questions (consider landmark data acquired from contact digitisers: this data samples only gross morphology, but is still useful for many forms of statistical hypothesis testing, even without surfacing). If, on the other hand, we wish to study the morphology of small structures, such as the semicircular canals of the inner ear, then even standard medical CT may not sample the specimen at a sufficient density.

To consider how sparsity of data effects the surface reconstruction, let us imagine a data set that consists of point measurements from a plane. If we are using a linear surface representation, the reconstruction is exactly equivalent to determining the original plane. Indeed, this constraint means that we need very few point measurements. We will offer, without proof, the fact that only three non-collinear surface measurements are needed. Using these three measurements, our surface representation will exactly match the plane; we are, in fact, missing no data at all.

For the case of a plane, or most other basic geometric primitives such as the conic sections, the construction of a surface representation is often fairly trivial, both in terms of the number of required measurements to “accurately” (and in this case, but not generally, *exactly*) perform the reconstruction, and in terms of processing. Unfortunately, real world fossil specimens are rarely ever as uniform in shape as a plane, or even a sphere: bone typically offers a surface with varying curvature, and all measurements have some levels of random noise associated with them. Uniform curves can be reconstructed using few point measurements, such as with the previously mentioned conic sections. But as the curvature begins to vary (as with bony surfaces) and the surface is no longer some idealised form (e.g., a conic section), more measurements are needed in its reconstruction. We can obtain an idea of how densely the surface should be sampled by considering the problem in terms of signal processing, which we do in section 4.2.2. For now, we move on to discussing the somewhat more trivial matter of determining isosurfaces from volume data.

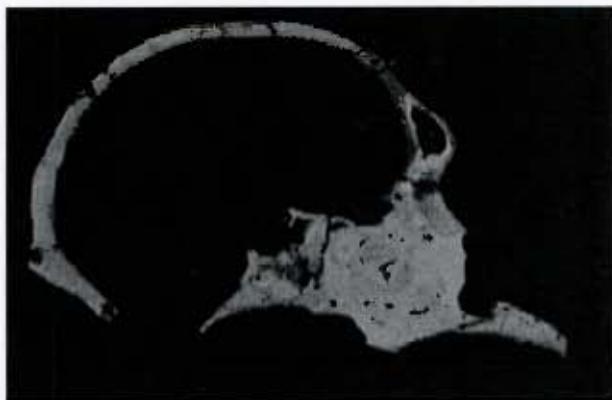


Figure 9: A mid-sagittal CT slice through STS 5.

#### 4.2.1 Processing of Volumetric Data

A common algorithm for initially processing volumetric data is Lorensen and Cline's (1987) *Marching cubes* algorithm. As input, marching cubes takes image slices from CT or other radiological modalities supplying volumetric information, as well as a density / isovalue supplied by the user. The original paper (Lorensen and Cline, 1987) applies the technique to CT, MRI and *Single Photon Emission Computed Tomography* (SPECT) data. Let us call the user supplied isovalue  $\delta$ . This value represents the density of the isosurface to be extracted by the algorithm. In figure 9 we can see a mid-sagittal section of STS 5 obtained using CT; higher density material is represented by the lighter areas in the scan, while cavities — areas of no density — are black. Figure 10 on the next page is a corresponding isosurface produced from a set of slices such as that in figure 9.

Marching cubes arranges the CT or other slices on top of each other in a “stack”. For each pair of slices the algorithm examines eight adjacent pixels, four from each slice. These eight pixels form a cube, a so called *voxel*. The algorithm processes all such cubes; it could be said that it *marches* across the cubes.

Now, each of the eight vertices of the voxel contain density information taken from one of the original slices. Some of these vertices will have a density value below that supplied by the user, while others a value above; the algorithm classifies the vertices accordingly. We know that the isosurface intersects the edges which have one vertex of greater density than  $\delta$ , and the other below  $\delta$ . Because the voxel cube has eight vertices, there exists  $2^8 = 256$  possible intersection patterns between the isosurface and the voxel. The authors reduce this number to fourteen by using the voxel's rotational symmetry (it is, after all, only a cube), and complimentary cases: for example, the isosurface produced when only one vertex is above the  $\delta$  value is exactly the same as the isosurface produced if that vertex were below the  $\delta$  threshold, and all others above. The first step



Figure 10: This is a surface model of STS 5, produced from CT data such as that in image 9, using the marching cubes algorithm. Take note of the rippling just behind the supraorbital ridges: this is artifacting produced by a helical CT scanner. The cut in the neurocranium is not an artifact: it was the result of limestone blasting that occurred at the Sterkfontein caves which cleanly cut the calvaria from the cranium.

in the algorithm is to classify all vertices as lying either above or below  $\delta$ , and then select from the fourteen combinations to determine the shape of the isosurface in the voxel cell.

The shape of the isosurface in each voxel tells us which edges the isosurface intersects; the algorithm's next step is to determine the exact intersection point. This is done by interpolating between the density values of the edge vertices to find the position on the edge with density value  $\delta$ . This is the intersection point.

There are a number of special cases to consider. The two most important occur when the density value of every vertex in a given voxel is above or below the supplied  $\delta$  value. This indicates that the isosurface defined by  $\delta$  does not intersect the voxel: the voxel lies either in the "interior" of the isosurface (all density values are above  $\delta$ ), or the "exterior" (all values are below). Such voxels need no further processing.

The marching cubes algorithm has been extended, by, among others, Wilhelms and Gelder (1990) and Ning and Bloomenthal (1993). This has principally been to remove ambiguities in the original characterisation of isosurface shape using the 256 intersection combinations.

A number of other techniques exist for processing and visualising volumetric data. For example, ray casting techniques use a line cast through the voxel grid to search for the intersection with

the isosurface, allowing for isosurface determination. Other techniques, such as Levoy (1990), Brady et al. (1997) and Roettger et al. (2000) attempt to render the volumetric data directly. We have, however, used marching cubes to produce the surface representation we have used in our work, and so will not be discussing the other techniques in any detail. Figure 10 shows a surface produced for STS 5.

#### 4.2.2 Processing Surface Data

There has been much research interest in constructing surface models from point cloud data. As Hoppe et al. (1992) mention, much of this work has focused on special cases. For example, a laser scanner's point measurements lie on a regular grid. With a few caveats it is fairly trivial to turn this regular grid into a surface model by connecting adjacent points in a regular pattern. A laser scanner, however, usually needs more than one scan to completely measure an object — since portions of the object will fall within the “shadow” of the laser from any given position. These multiple scans need to be combined to create a surface model. Methods to do so exist, such as Turk and Levoy (1994), which constructs a separate surface for each scan which is then stitched together to form a complete surface. This is clearly an example of a special case, and the technique is inappropriate for surface reconstruction of an arbitrary point cloud (i.e., where the points do not lie in a regular grid as with laser scan data).

The data provided by photogrammetry is, unfortunately, not sufficiently regular enough to be amenable to the method of Turk and Levoy (1994). More general methods are needed for this data, such as that of DeRose et al. (1992) and Hoppe et al. (1992, 1993, 1994), appropriately called *surface reconstruction from unorganised points*. Although the surface reconstruction field had been in existence for some time before this work — even within computer graphics — Hoppe et al.'s work is sometimes viewed as introducing the field to computer graphics (see, for example, Amenta and Bern (1998)), but consider Boissonnat (1984) or Pratt (1987) as examples of earlier work within the computer graphics field.

Hoppe et al.'s technique is approximative, rather than interpolative. It assumes that the input data is noisy (usually a valid assumption) and that smoothing of the data will not lower the quality of the reconstruction. Their approach first approximates the surface by placing tangent planes at each of the sample points. These tangent planes are used determine a signed distance function  $f$  that estimates the distance from the sample points to the desired surface. Then, the set of all points that take  $f$  to zero (i.e.,  $\{x | f(x) = 0\}$ , the so called *zero set* of  $f$ ) is the required surface. In light of the previous section, the reader might find it interesting that Hoppe et al. use marching cubes to trace the isosurface defined by the zero set. The original technique produces a linear surface model, but this has been extended to produce higher order models (Hoppe et al., 1994).

Zero set based reconstructions are but one family of techniques used in the surface reconstruction field. A tool common to other families is the *Delauney triangulation* of a sample set, and its dual, the *Voronoi diagram* of the set. These techniques produce interpolating — and not approximating — surfaces. They are spatial decompositions, and can best be illustrated by a diagram; the reader should consult figures 12(a) and 12(b) while reading the following descriptions.

Let our sample set,  $S \subset \mathbb{R}^n$  be a set of measurements from some space  $\mathbb{R}^n$ , usually either  $\mathbb{R}^2$  or  $\mathbb{R}^3$ . Let us consider a single sample point,  $s \in S$ . The *Voronoi cell* of  $s$  is the subset of  $\mathbb{R}^n$  that is closer to  $s$  than any other sample point. The decomposition of  $\mathbb{R}^n$  into the set of Voronoi cells corresponding to each sample point in  $S$  is the *Voronoi diagram* of  $S$ . Each Voronoi cell defines a polytope — a generalisation of a polygon to higher dimensions — that separates it from the surrounding Voronoi cells. The vertices of these polytopes are the *Voronoi vertices*. It can be shown that each vertex will be equidistant from  $n+1$  sample points in  $S$  (i.e.,  $3+1 = 4$  points in  $\mathbb{R}^3$  and  $2+1 = 3$  points in  $\mathbb{R}^2$ ). These  $n+1$  points define a *Delauney simplex* (a triangle in 2D, figure 11(a), a tetrahedron in 3D, figure 11(b), and the set of these simplices is the *Delauney Triangulation* of  $S$ . For an introduction to the mathematics of Delauney triangulations and Voronoi diagrams the reader may consult any computational geometry text book, such as Preparata and Shamos (1985).

Now, it might have occurred to the reader that a Delauney simplex is always a structure of one too many dimensions than actually required. A surface model in a 3D Euclidean space is a 2D structure but a Delauney triangulation in a 3D Euclidean space produces a 3D structure: a tetrahedral lattice (figure 11(b)). This generalises to Euclidean spaces of any dimension. Boissonnat (1984) — who proposed the use of Delauney triangulation for surface reconstruction — likens the Delauney triangulation to a *volumetric reconstruction*<sup>8</sup> of an object: only some of the faces of the Delauney simplex lie on the surface of the object, the others, and indeed the simplex's interior itself, lie on the *interior* of the object. To create a surface from a Delauney triangulation, one has to remove the faces from each simplex that do not lie on the surface.

A technique of some prominence that is based on “thinning” Delauney triangulations is the  $\alpha$ -*shapes* method (Edelsbrunner et al., 1983; Edelsbrunner and Mücke, 1994). Given a sample set,  $S \subset \mathbb{R}^3$ , the  $\alpha$ -shapes are a family of surfaces with each individual surface defined by a value of  $\alpha$ ,  $0 \leq \alpha \leq \infty$ . Note that  $\alpha$  is free to take on the value of infinity: in which case the  $\alpha$  shape is then exactly identical to the *convex hull* of the samples  $S$ . The convex hull is the smallest convex



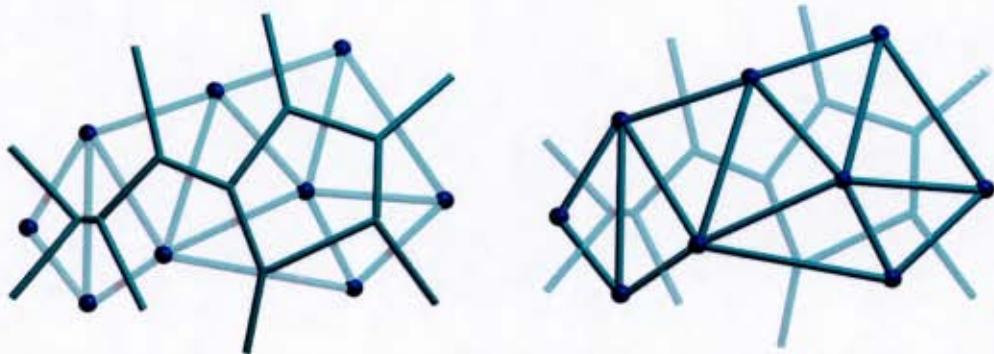
(a) A simplex in 2D: the blue points are sample points. Note that it is simply a triangle.



(b) A simplex in 3D: the blue points are sample points. This is essentially the triangle in figure 11(a), but with one extra point.

Figure 11: Simplices in 2D and 3D.

<sup>8</sup>Consider the relationship between volume and surface: volume has one more dimension than the surface of an object.



(a) The darker lines show Voronoi diagram of the points. The diagram partitions the spaces into the regions that are closest to each point.  
 (b) The darker lines show the Delauney triangulation of the points.

Figure 12: The Voronoi diagram and its dual, the Delauney triangulation.

polytope containing the sample points  $S$ . As its name suggests, the convex hull cannot contain any cavities or holes. For  $\alpha = 0$ , the  $\alpha$ -shape is the set  $S$  itself. All the “interesting” surfaces that we may wish to use in the reconstruction have  $\alpha$  values between these two extremes. As  $\alpha$  decreases from infinity the surface can develop holes and cavities, until the surface disappears at  $\alpha = 0$ . The authors liken  $\alpha$ -shapes to filling  $\mathbb{R}^3$  with Styrofoam and representing each sample point in  $S$  with a rock (Edelsbrunner and Mücke, 1994). The  $\alpha$ -shape algorithm is akin to using an eraser of diameter  $\alpha$  to remove as much Styrofoam as possible without upsetting the rocks. The remaining Styrofoam defines the  $\alpha$ -shape. The  $\alpha$ -shapes technique is related to the Delauney Triangulation in that the family of  $\alpha$ -shapes can be shown to be represented by a Delauney triangulation. Calculating a particular  $\alpha$ -shape is then equivalent to thinning the Delauney triangulation by choosing appropriate faces from each simplex so as to reveal the desired  $\alpha$ -shape.

One possible critique of the  $\alpha$ -shapes technique is its reliance on the  $\alpha$  parameter. The user is unable to simply perform the surface reconstruction using the technique, but must instead explore the space of possible surfaces by varying the value of  $\alpha$  and finally settling on a reconstruction that she feels is adequate for her needs.

A more user-friendly technique is that of Amenta et al. (1998) and Amenta and Bern (1998). This calculates a surface model (which the authors call the *crust*) using the Delauney triangulation. Instead of calculating the Delauney triangulation of  $S$  directly, the crust algorithm calculates the Delauney triangulation on  $S \cup M$ , where  $M$  is a set of points that approximate the *medial axis* of the surface under construction. The medial axis (see figure 14) of a surface is the set of all points for which a sphere centred at the point intersects the surface twice or more. These sphere centres are those points which have more than one closest point on the surface. They can be thought of as



Figure 13: These are surface reconstructions of surface data from STS 5. The surface data was obtained from the photogrammetric method outlined in chapter 3. The surface reconstruction was performed using (Amenta et al., 1998). Although we have captured certain surface detail, there are areas suffering from a lack of sampling density, most notably in the basicranium — a “scarcity” of data, as spoken of in the text. The user might find it illuminating to compare the results obtained through the surface reconstruction from photogrammetric surface data with that obtained from the volumetric data, shown in figure 10.

laying along the object's "centre"<sup>9</sup>. The distance between a point on the medial axis and its closest surface points is a proxy for surface curvature: as the curvature increases, the medial axis moves closer to the surface. At sharp corners — at *creases* — the medial axis touches the surface.

Amenta et al. (1998) use a subset of the Voronoi vertices to approximate the medial axis: in other words,  $M \subset V$ . The thinning of the Delaunay triangulation follows simply: the surface model consists of the faces of the Delaunay simplices whose vertices are all in  $S$ . The Voronoi vertices  $M$  approximating the medial axis are used to *filter* the faces of the triangulation, and the authors call the technique *Voronoi filtering* (Amenta and Bern, 1998). The algorithm's fidelity relies on the sample  $S$  being dense enough that each simplex in the Delaunay triangulation of  $S \cup M$  will always contain one vertex in  $M$ . Figure 13 shows a Voronoi filtered surface reconstruction of STS 5 from point clouds obtained using the photogrammetric method of Smit (1997) described in chapter 3.

We can now discuss what we mean by "scarcity" of data, and a "dense enough" sampling. Many of the Delaunay based surface reconstruction algorithms have a sampling criterion stating how dense the sample set should be in order to create an accurate reconstruction of the surface. The technique of Amenta et al. is no different. Specifically, given a surface  $F$  and a set of points on the surface  $S$ , the authors define the set  $S$  to be an  $r$ -sample if the distance from any point  $p \in F$  to the nearest point in  $S$  is at most  $r$  times the distance from  $p$  to the closest point on the medial axis. An  $r$  value below one ensures that the sampling density is inversely proportional to the distance to the medial axis. Note that for any given value of  $r$ , the maximum distance between a surface point  $p$  and the closest sample point required for a set to remain an  $r$ -sample increases as the surface curvature decreases and the surface flattens out (and, of course, does not come closer to another surface). This results from the medial axis lying correspondingly further from the surface as the curvature decreases. The authors prove several theorems concerning the surface reconstruction for values of  $r$  less than 0.06, but note that the algorithm works well for values less than 0.5. Although Amenta et al. (1998) provide no reason for this, we can gain some insight into why this is so from sampling theory.

A surface can be thought of as a continuous *signal*. Represented as such, changes in surface curvature could be thought of as changes in the signal's frequency. We can see that the signal could have arbitrary levels of "high-frequency" detail — indeed, corners and creases are areas with infinite

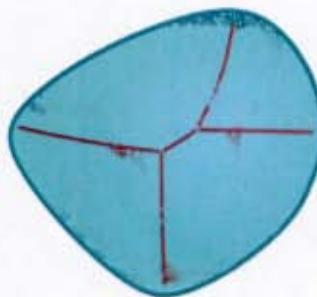


Figure 14: The red line shows the medial axis of the surface. Note that for this 2D structure, the medial axis is a line, or a 1D structure. It is always true that in a  $d$  dimensional space, the medial axis is a  $d - 1$  dimensional structure.

<sup>9</sup>The medial axis is, interestingly, a biological concept, introduced by Blum (1973).

frequency (Foley et al., 1996)<sup>10</sup>. If we sample this surface, the sample set is clearly a discrete set (since we can only measure a finite number of samples). Because there is a finite set of measurements, we are unable to measure changes in a surface that contains areas of too high a frequency (analogous with the 2D case). Because we have failed to sufficiently sample the surface, we are unable to perfectly reconstruct the surface. The case is not as bleak as it may sound, since in analogy to the *Nyquist rate* from signal processing, we can create surface models when our sampling rate is greater than twice the “frequency” of the surface being sampled. There are a number of books that the reader may consult concerning signal theory. As always, Foley et al. (1996) offers a basic introduction to 2D signal processing theory in terms of computer graphics. Gonzalez and Wintz (1987) provide a more detailed introduction to digital image processing.

It should be noted that volumetric data is itself effected by the Nyquist sampling rate, since volumetric data is essentially a scalar field of density samples.

In our work, we use Amenta et al.’s algorithm for surfacing data obtained through the photogrammetric technique outlined in chapter 3. Figure 13 shows the surface obtained from the STS 5 photogrammetric data. The preferred surface reconstruction was, however, obtained from CT data with the marching cubes algorithm — as mentioned in chapter 3, the CT data suffers from less noise and has a denser sampling than the photogrammetric data. This resulted in a higher quality surface model (as is obvious from comparing figures 10 and 13).

### 4.3 Rendering

Once surface models have been constructed, standard rendering algorithms can be used to display the models to the user. We will not go into details concerning these basic algorithms here; the reader can consult any introductory text on computer graphics, while chapter five of Zollikofer and Ponce de León’s book on virtual reconstruction of fossil material (Zollikofer and Ponce de León, 2005) also contains some introductory information on displaying surface models.

---

<sup>10</sup>This discussion is taken from 2D signal processing. It is trivial to show that height maps are equivalent to the 2D signal domain. What is less trivial to demonstrate is that every 2-manifold model can be partitioned into a sequence of height maps. We can see this by noting that height maps are homeomorphic to a disc: they are bordered and have no overlapping areas. The technique of Lévy et al. (2002) can be used to partition a surface representation into a set of patches homeomorphic to a disc, and the discussion concerning sampling theory then applies to these portions.

### 4.3.1 Rendering Point Clouds

There are many techniques used to render three dimensional objects. Most techniques concern themselves with the rendering of surface models. There are, however, a growing number of algorithms that concern themselves with the rendering not of surface models directly, but of point clouds that are rendered so as to give the appearance of a continuous surface.

The use of points as a basic rendering primitive was first proposed by Levoy and Whitted (1985). Levoy and Whitted feel that the representation of surfaces as points for the purpose of rendering has advantages over the use of conventional rendering primitives such as polygonal surfaces, implicit surfaces, and so on, namely:

- There is no need to adapt existing rendering algorithms to display a new form of surface primitive. They proposed the use of points as a *meta-primitive* to which all other geometric primitives would be converted for rendering, and which would allow the use of a standard rendering pipeline. No updating of the pipeline is needed for the introduction of new modelling primitive, only a way to convert the new primitive into a set of points. However, over the years the meta-primitive of choice has become not the point but the triangle.
- As the complexity of the object increases to the point where each primitive projects to less than a pixel in screen space, the use of primitives larger than a point in order to gain performance advantages from coherence (part of the reason for using triangles) becomes counter-productive. Although most models have not reached this level of complexity and detail, the introduction of better data acquisition devices — such as laser range scanners and microCT — have produced data volumes where this level of density is achievable.

Levoy and Whitted's basic approach is to take each pixel in the resulting image and calculate the distance from the pixel to each point in image space. This distance is then weighted by a Gaussian function centred around the pixel, and this value is the contribution of the point to the value of the pixel. Since each point contributes to each pixel, it allows for the point cloud to appear solid when rendered.

Other authors have extended this approach, for example:

- *QSplat* (Rusinkiewicz and Levoy, 2000) A performance point rendering method. Their technique combines multiresolution model representations, culling and level of detail in order to render extremely large models at interactive rates.
- *Surfels* (Pfister et al., 2000) This technique is closely related to the work of Levoy and Whitted (1985). Surfels are “surface elements”, essentially points, which contain further information concerning the object, such as normals and a texture sample.

- *Surface Splatting* (Zwicker et al., 2001) This work extends the texturing formulation of Heckbert (1989) to point rendered geometries in order to obtain high quality texturing effects for point rendered surfaces.

Point rendering techniques are ultimately useful as a means to simplify and accelerate model rendering — especially that of large models — by replacing triangles as the basic primitive of choice. These techniques, though, are not necessarily useful as a means of directly rendering acquired point clouds and thereby avoiding the need for surface reconstruction. Consider the Rusinkiewicz and Levoy (2000) QSplat technique, which requires the connectivity information supplied by a triangle mesh in order to process the model so as to render the surface without any holes.

#### 4.4 Closing Remarks

The work carried out in this thesis made use of a number of the techniques presented in this chapter. Most notably, we made use of the marching cubes algorithm, and Amenta et al.'s crust algorithm. Results of the algorithms can be seen in figure 10, which is a surface model of STS 5 acquired from CT data using the marching cubes algorithm. Figure 13 on page 56 shows a surface representation of STS 5 obtained through photogrammetry, and the application of the crust algorithm.

# Chapter 5

## The Correction Techniques

This chapter describes the three techniques used in this thesis for estimating missing landmarks: mean substitution, thin plate splines and multiple linear regression. It begins by discussing various aspects and methods that the reader should be aware of (such as the problem of sample sizes; landmarks vs semilandmarks), then moves on to the techniques themselves. This chapter builds upon section 2.6, which discusses some basic morphometric concepts.

### 5.1 Aspects of Statistical and Geometric Reconstruction

Analytical reconstruction techniques proceed in a twofold manner:

1. Missing landmarks are estimated
2. Interlandmark morphology is corrected.

Most morphometric analyses concern themselves solely with landmarks to the exclusion of interlandmark data. Because of this, much of the reconstruction work has stopped at the landmark (e.g., Richtsmeier et al., 1992) or semilandmark (e.g., Gunz, 2005) level of morphological detail. Reconstructions dealing with interlandmark morphology have often been performed manually (e.g., most of the Zollikofer et al., 2005, reconstruction).

There is, however, a growing trend to use geometric approaches to interlandmark reconstruction. These have been carried out preeminently through the use of thin plate splines, mostly in the work of Zollikofer and Ponce de León. Its use for both landmark and interlandmark correction is thus covered in two separate sections of this chapter, namely sections 5.2.2 on page 78 — concerning landmark correction — and section 5.2.5 on page 87 — which covers the correction of the remaining interlandmark morphology.

First, though, a discussion concerning landmarks and semilandmarks is presented.

### 5.1.1 Landmarks and Semilandmarks

Landmarks are usually considered to be “missing” if they are in some way related to damaged morphology. These landmarks may be completely missing, as when portions of a fossil specimen are fragmented and lost; unidentifiable, as when the associated morphology of the fossil specimen has suffered extreme plastic distortion; or identifiable but out of their anatomical relationships with the remaining morphology, and hence meaningless. Missing landmarks are estimated using some form of statistical (e.g., mean substitution, multiple linear regression) or geometrical (e.g., thin plate splines) method.

While landmark based analyses are common, not all morphology is readily amenable to landmark based morphometric methods. Consider landmarks that attempt to capture morphology not easily defined by points: the morphology of curves and surfaces (*type 3 landmarks* in Bookstein’s (1991) classification of landmarks). These landmarks are often points of maximum curvature, maximum length, and so on, and do not have a clearly meaningful homologous point on another individual. Because of this lack of homology they are considered *deficient* as landmarks, and their use in standard landmark based morphometric analyses is questionable.

Unfortunately, there are many anatomical areas which have a poor collection of non-deficient landmarks, the brain case being a prime example. Here there are relatively few structures — such as the meeting points of sutures, or bony processes — on which landmarks may be situated. If landmark methods are to be used to their full utility, curve and surface data must somehow be made useful for landmark based analyses. The method of semilandmarks is such an attempt.

Semilandmarks are a method to operationalise curves and surfaces in terms of landmarks. The main obstacle is the lack of homology between arbitrary points along a curve or surface. While biological homology is defined in terms of evolutionary history and common descent, morphometrics needs to define the homology between two individuals in a practical way amendable to quantitative analysis. This is currently done using interpolation methods such as thin plate splines (Bookstein, 1991). Using thin plate splines, semilandmarks attempt to minimise any undue homology transformations between two individuals caused by misplaced type 3 landmarks. This is done by shifting the *semilandmarks* — the type 3 landmarks — along the surface of one of the two specimens related via thin plate spline homology map, thereby changing the homology map defined by the thin plate spline. The semilandmark regimen assumes that if the semilandmarks can be shifted to create a thin plate spline with lower bending energy (as measured by equation (1) in 2D) — and hence with a smaller amount of transformation between the two landmark configurations — then this is the preferred system of splines. Intuitively, this seems an attractive approach: a thin plate spline system

with greater bending energy than another warps the space between the two forms to a greater degree than a spline system with lesser bending energy. These “high energy” warps are often characterised by unwanted twisting and shearing between forms, and in the worst case, foldover effects<sup>1</sup>. The semilandmark method attempts to define a thin plate spline homology map which lacks such high energy warping effects by repositioning the type 3 landmarks, on the assumption that this is always the best approach when operationalising these deficient landmarks.

Semilandmarks may be produced automatically, by randomly scattering them in rough correspondence over two individuals. The position of the semilandmarks are then shifted to produce a satisfactory spline as a homology map. This allows for a large number of semilandmarks to be used within an analysis, far more than there are identifiable landmarks.

The positions of semilandmarks are calculated with respect to the true landmarks. This restricts the semilandmarks’ movement, since some of the thin plate spline’s bending energy will have been contributed by the true landmarks themselves. The landmarks are not shifted during the positioning of the semilandmarks, which means that the method cannot drive the bending energy of the thin plate spline completely towards zero, a case which would make for an uninteresting homology map.

Semilandmarks, while mentioned in the appendix of Bookstein (1991), are more properly introduced in Bookstein (1997). While Bookstein was more concerned with the semilandmarks in  $\mathbb{R}^2$ , he does briefly outline the  $\mathbb{R}^3$  case (Bookstein, 1997), which is further described by his student in Gunz (2005) and Gunz et al. (2005).

While the method of semilandmarks is not used in the work presented here, others have begun to use the regime both in reconstructive work (e.g., Gunz, 2005), and in morphometric analyses (e.g., Bookstein et al., 2003; Martinón-Torres et al., 2006). To this end, they have been introduced here to allow a discussion of, and comparison to, other work.

### 5.1.2 Sample Sizes

Perhaps the main concern when performing statistical analyses on fossil material is sample sizes — and when we consider the australopith fossil collection, especially when broken down by species, these sample sizes can be small indeed.

For example, *Sahelanthropus tchadensis* (Brunet et al., 2002) has less than ten discovered specimens, only one of which is cranial, and this specimen is unfortunately damaged. A similar lack is true for species such as *Orrorin tugenensis* (Senut et al., 2001) and *Ardipithecus ramidus* (White et al., 1994). This problem is compounded by the relatively recent discovery of this material. Others, such as *Australopithecus africanus* or *Australopithecus afarensis*, have had decades of work in which

---

<sup>1</sup>Foldover effects may result in portions of anatomy overlapping and intersecting one another.

material has been collected and described. Further, especially in the case of *O. tugenensis* and *A. ramidus*, published description concerning existing material can be meagre, even considering that *A. ramidus* was originally described in 1994 (under the name *Australopithecus ramidus*). This again limits our reasoning about these species.

Even when a species has a large body of preserved material, this does not imply that the specimens are *well* preserved. *Australopithecus africanus*, with a large collection of fossil remains, has very few well preserved crania. Wood and Richmond (2000) note four:

- Taung 1 (Dart's Taung Child).
- Sts 5 (Mrs. Ples)
- Sts 71
- Stw 505

Concerning *Paranthropus robustus*, Wood and Richmond (2000) note only one well preserved cranium, SK 48, which nonetheless suffers from some plastic distortion. Of the listed *A. africanus* specimens, Sts 5 and Taung 1 are the least damaged specimens, although Taung 1 is juvenile. Sts 71 and Stw 505 are both missing anatomy and suffer from some mild plastic deformation.

Another — more practical — limitation on sample size is the understandable difficulty in obtaining access, whether physically or electronically, to some of these specimens. For a discussion on this, along with a suggestion concerning a digital archive for models of fossil specimens, see Weber (2001).

These small sample sizes limit our reasoning concerning the fossil material. For instance, Smith (2005) notes the difficulty in determining a new specimen's taxonomic affinity when the sample sizes of known and related species are small, and how these small sample sizes can effect the interpretation of this specimen's anatomical details (to paraphrase: when you have not seen many individuals, anatomy never before seen — but not outside the species' natural variation — appears unique and can unduly colour the interpretation of a new find).

This lack of available fossil material is evident in the morphometric analyses and various reconstructions that have been performed, where typical reference sample sizes used to perform estimation of missing landmarks and interlandmark morphology are often trivially small, not infrequently in the order of one individual (e.g., Gunz, 2005; Ponce De León and Zollikofer, 1999; Zollikofer and Ponce de León, 1995, 2005). The problem is essentially identical to that discussed by Smith (2005): if its morphology is known only from one or two other individuals (damaged or otherwise), how can its morphology be reconstructed?

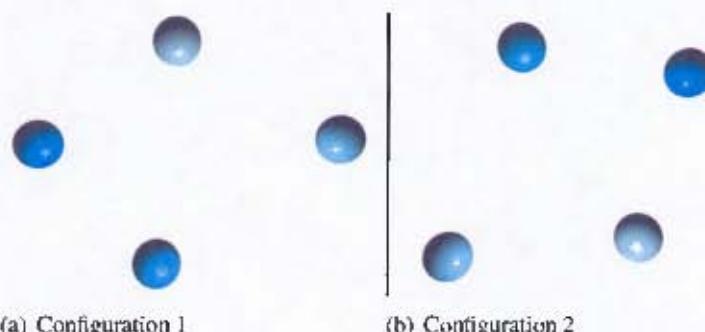


Figure 15: This figure demonstrates the rotation independence of multidimensional scaling. Given the distances between the given points, both configuration 1 and 2 are possible outcomes of running MDS. This is because both configurations of points have the same distances between the points. Similarly, the configurations as a whole could have different translations in space: because the points still retain the same distance between them, each configuration would still be a correct output of the MDS method.

While work has been performed on the relative relationships in variance / covariance patterns between primate species (as with Ackermann, 2002, 2005; Ackermann and Cheverud, 2000, 2002), building reference samples from species other than that of the individual under reconstruction has yet to be widely applied. There have, however, been a few exceptions, such as the fitting of modern human endocranial anatomy onto a Neandertal cranium by Ponce De León and Zollikofer (1999). While many studies indicate that extant primate species share variance / covariance structures, no work has yet been done on taking advantage of this, in, for example, regression analysis.

The work in the following chapters investigates how using members of another species as a reference sample impacts the outcome of fossil reconstruction, with test results in chapter 6, and a discussion in chapter 7.

### 5.1.3 Multidimensional Scaling

Multidimensional scaling (MDS) is a set of techniques useful for visualising multivariate data (Cox and Cox, 1994). Its core consists of methods that match points in a low dimensional Euclidean space (i.e.,  $\mathbb{R}^2$  or  $\mathbb{R}^3$ ) to objects (perhaps of unknown dimension) for which the only known information is their proximity to one another (Cox and Cox, 1994). The distances between the resulting points are the same as (or related to in some well defined way) the original proximities between the objects.

The usual example given to demonstrate the use of MDS involves the distances between cities on a map. Using only these distances, MDS can calculate the relative position of each city (using the technique called *classical scaling*, described below). However, the solution is not unique: the points are only correctly positioned *relative* to each other. The actual translation, rotation and reflection

of the set of points as a whole is completely arbitrary: under any rigid body transformation the distances between the points in the configuration remain the same, and hence any of the above transformations of the points remain a valid solution to the multidimensional scaling problem, as demonstrated in figure 15 (Cox and Cox, 1994).

Multidimensional scaling is useful to techniques based on Euclidean Distance Matrix Analysis (EDMA — see chapter 2), where morphology is not represented directly as a sequence of landmarks, but as a matrix of distances. MDS allows distance matrices to be mapped back to landmarks, and is used in this way by the multiple linear regression correction method, as described later in the chapter (section 5.2.3).

The input data to the various MDS techniques is typically some form of *proximity* information:

- *Similarities.* A similarity of 0 shows that two objects are not similar at all. The higher the value (up to some predefined maximum) the more similar the objects are. At the maximum they are considered identical.
- *Dissimilarities.* Standard distances are an example of dissimilarities: identical points have a dissimilarity of 0. The higher the dissimilarity, the more “dissimilar” the points.

MDS techniques differ primarily in how they match the distances of the calculated points to the original proximity data (Cox and Cox, 1994). For example, if it can be assumed that the available proximity data is exactly the Euclidean distance between the objects, *classical scaling* (described below) may be used to find Euclidean points representing the objects. *Least squares scaling*, on the other hand, assumes that the given proximity data will not be the same as the reconstructed distances. A continuous, monotonic function is used to map these proximities to the distances. This function is determined by the least squares minimisation of an optimisation function on the distance between the actual Euclidean distance values and the object proximities as transformed by the function.

### Classical Scaling

Classical scaling assumes that the proximities are exactly the Euclidean distances between the points calculated by the method. This is the standard problem domain when transforming Euclidean distance matrices into landmarks.

The input to the technique is an  $n \times n$  matrix of squared distances,  $(d_{ij}^2)$ , where

$$d_{ij}^2 = (x_i - x_j)^T(x_i - x_j) \quad (2)$$

is the squared distance between the unknown points  $x_i$  and  $x_j$ . The points  $x_i$ ,  $1 \leq i \leq n$  are all  $p$  dimensional, although in reality this dimension may be unknown, and only estimated. With the input of squared distances, classical scaling discovers the values of all the  $x_i$ ; the distances between the  $x_i$  are as specified by the various  $d_{ij}$  values.

Classical scaling may be performed in the following way:

First, assume that the centre of mass of all the  $x_i$  is the origin (i.e.,  $\frac{1}{n} \sum_{i=1}^n x_i = 0$ ). This serves to remove the effect of arbitrary translations on the points<sup>2</sup>.

The key to calculating the points is to construct the matrix  $B = (b_{rs})$ , where

$$b_{rs} = x_r^T x_s \quad (3)$$

Since each  $x_i$  is unknown,  $B$  cannot be directly calculated; instead, we recast  $B$  in terms of  $(d_{ij}^2)$ .

Performing a “quadratic expansion” of equation 2, we see that

$$d_{ij}^2 = x_i^T x_i + x_j^T x_j - 2x_i^T x_j$$

From this we form the following sums:

$$\frac{1}{n} \sum_{i=1}^n d_{ij}^2 = \frac{1}{n} \sum_{i=1}^n x_i^T x_i + x_j^T x_j \quad (4)$$

$$\frac{1}{n} \sum_{j=1}^n d_{ij}^2 = x_i^T x_i + \frac{1}{n} \sum_{j=1}^n x_j^T x_j \quad (5)$$

Note that the terms which have vanished in each of the above equations have done so using the fact that the centre of mass is zero, causing terms of the form  $-\frac{1}{n} \sum_{i=1}^n 2x_i^T x_j$  to evaluate to 0.

Now, keeping equations 4 to 5 in mind:

$$\begin{aligned} d_{ij}^2 &= x_i^T x_i + x_j^T x_j - 2x_i^T x_j \\ &= \frac{1}{n} \sum_{j=1}^n d_{ij}^2 - \frac{1}{n} \sum_{j=1}^n x_j^T x_j \\ &\quad + \frac{1}{n} \sum_{i=1}^n d_{ij}^2 - \frac{1}{n} \sum_{i=1}^n x_i^T x_i \\ &\quad - 2x_i^T x_j \end{aligned}$$

---

<sup>2</sup>This is not as restrictive an assumption as it sounds. We can always replace the subsequent  $B$  matrix with  $B' = HBH$ , where  $H$  is the centring matrix  $H = I - \frac{1}{n}\vec{1}\cdot\vec{1}^T$ . Here,  $\vec{1}$  is the column vector of  $n$  ones, and  $I$  is the  $n$  dimensional identity matrix. This serves the same effect as the above assumption

$$\begin{aligned}
&= \frac{1}{n} \sum_{j=1}^n d_{ij}^2 - \frac{1}{n} \sum_{j=1}^n \left( \frac{1}{n} \sum_{i=1}^n d_{ij}^2 - \frac{1}{n} \sum_{i=1}^n x_i^T x_i \right) \\
&\quad + \frac{1}{n} \sum_{i=1}^n d_{ij}^2 - \frac{1}{n} \sum_{i=1}^n x_i^T x_i \\
&\quad - 2x_i^T x_j \\
&= \frac{1}{n} \sum_{j=1}^n d_{ij}^2 - \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n d_{ij}^2 + \frac{1}{n} \sum_{i=1}^n x_i^T x_i \\
&\quad + \frac{1}{n} \sum_{i=1}^n d_{ij}^2 - \frac{1}{n} \sum_{i=1}^n x_i^T x_i \\
&\quad - 2x_i^T x_j \\
\Rightarrow x_i^T x_j &= -\frac{1}{2} (d_{ij}^2 - \frac{1}{n} \sum_{i=1}^n d_{ij}^2 \\
&\quad - \frac{1}{n} \sum_{j=1}^n d_{ij}^2 + \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n d_{ij}^2)
\end{aligned}$$

$B$  has now been rewritten in terms of the squared distance matrix, rather than explicitly in terms of the  $x_i$ s. After Cox and Cox (1994), let  $a_{ij} = -\frac{1}{2}d_{ij}^2$ , then form the sums

$$a_{i\cdot} = \frac{1}{n} \sum_{j=1}^n a_{ij} \tag{6}$$

$$a_{\cdot j} = \frac{1}{n} \sum_{i=1}^n a_{ij} \tag{7}$$

$$a_{..} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \tag{8}$$

$B$  may now be written more succinctly as

$$b_{ij} = x_i^T x_j = a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{..} \tag{9}$$

From equation 3,  $B$  can also be written as

$$B = X X^T \tag{10}$$

where  $X$  is an  $n \times p$  matrix containing the  $n$  points, a single point on each row. The matrix  $X$  is an unknown, and can be solved by considering the spectral decomposition of  $B$  in equation 9:

$$B = V \Lambda V^T \tag{11}$$

$\Lambda$  is a diagonal matrix carrying the eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$  of  $B$ , and  $V = [v_1, \dots, v_n]$  contains the respective eigenvectors,  $v_i$ . Assume that the eigenvalues are sorted, so that the matrix component

$\Lambda_{11}$  is the largest eigenvalue, and these decrease down the diagonal to  $\Lambda_{nn}$ , which contains the smallest eigenvalue. Also, assume that the eigenvectors are normalised.

Recall that the true dimension of the vectors  $x_i$  is possibly unknown. If one wishes to construct the points  $x_i$  in a  $q$  dimensional Euclidean space, they now need only take the  $q$  largest eigenvectors, creating  $V' = [v_1, \dots, v_q]$ . Each  $v_i$  should be scaled by the square root of their respective eigenvalues. If  $\Lambda' = \Lambda^{\frac{1}{2}}$ , we will show that:

$$X = V'\Lambda'$$

Let  $B' = V'\Lambda V'^T$ .  $B'$  is now a matrix whose role is equivalent to  $B$ , only with the number of eigenvectors and eigenvalues reduced to  $q$ . If  $q = p$  (as will be the case when classical scaling is employed in section 5.2.3), then there would be no reduction in the number of eigenvectors and eigenvalues, and  $B' = B$ . Now,

$$\begin{aligned} B' &= V'\Lambda V'^T \\ &= V'\Lambda'\Lambda'V'^T \\ &= V'\Lambda'(V'\Lambda')^T \\ &= X'X'^T \end{aligned}$$

$X'$  is an  $n \times q$  matrix of points in  $\mathbb{R}^q$ . If  $q = p$ , then  $X'$  will be exactly  $X$ . ■

### A Step by Step Approach to Classical Scaling

Once you have obtained the distance matrix, classical scaling may be performed as follows (Cox and Cox, 1994):

1. Calculate the  $a_{ij}$ , where  $a_{ij} = -\frac{1}{2}d_{ij}^2$ .
2. Calculate  $B = (b_{ij})$ , where  $b_{ij} = a_{ij} - a_{ii} - a_{jj} + a_{..}$  (equations 6 through 8).
3. Determine the eigenvalues and normalised eigenvectors of  $B$ . Scale the eigenvectors so that their length is equivalent to the square root of their respective eigenvalues.
4. Choose a dimension,  $q$ , in which you wish to reconstruct the points, and
5. then retain the  $q$  eigenvectors associated with the  $q$  largest eigenvalues. Ensure that these eigenvectors are sorted in decreasing order of their eigenvalues.
6. The points are then the rows of the  $n \times q$  matrix  $X'$ , where  $n$  is the number of points, and  $X' = [v_1, \dots, v_q]$ , where  $v_1, \dots, v_q$  are the scaled eigenvectors.

### 5.1.4 Registration

Registration is a process in which two sets of points,  $\{x_i\}$  and  $\{y_i\}$ , are brought into alignment. The sets are both of the same size, and contain corresponding points such that  $x_i$  corresponds to  $y_i$  for all  $i$ . Registration aligns the two sets by finding a rotation and translation (and possibly a scale factor) such that, when applied to the  $x_i$ , the distance between the  $x_i$  and  $y_i$  for all  $i$  is minimised. After registration, the points in the two sets lie as “close” as possible (Seeger and Labourey, 2000).

There have been many solutions to this problem. In Computer Science literature, the classic method is Horn’s quaternion based algorithm (Horn, 1987), in which the rotations are represented as quaternions (a generalisation of complex numbers). Other classic methods include those based on singular value decomposition (SVD), such as that presented by Cox and Cox (1994) and Seeger and Labourey (2000). SVD approaches allow for points in spaces of higher dimension than  $\mathbb{R}^3$  to be registered. Quaternion approaches, on the other hand, are restricted to  $\mathbb{R}^2$  and  $\mathbb{R}^3$ , although Horn (1987) feels that the quaternion approach is a superior solution because it will not perform any reflection, which SVD is liable to do; Besl and McKay (1992) feel that the quaternion approach is numerically more stable.

Registration methods using corresponding points have been extended to data sets in which exact correspondences may not be known. *Iterated Closest Points* (ICP) is just such an algorithm (Besl and McKay, 1992). The method assumes that the two sets lie in a rough, but not exact, alignment. The closest pairs of points between the two sets are then assumed to be corresponding, and a registration (using, for instance, Horn’s quaternion method, as the original ICP does) is performed. This is repeated over a number of iterations; on each iteration new corresponding points are chosen from the closest pairs of points. Once the position of the points between each iteration has changed less than some given value, the system is considered to have converged, and the process stops. ICP allows for the quick registration of data sets for which accurate corresponding points are unknown, such as in the registration of laser scan data from multiple stations.

In the applications presented here, the point correspondences are known, allowing for a direct solution to the registration problem. To this end, the SVD approach is described below. The method uses singular value decomposition of a matrix of cross covariances, and is based on the method of Cox and Cox (1994) and Seeger and Labourey (2000). The SVD approach has been chosen over a quaternion based method solely because quaternions are unable to represent reflections — in other words, changes in symmetry<sup>3</sup>. One set of points that requires registration will be obtained through classical scaling, which, as explained above, may produce points that have been reflected. Thus any

---

<sup>3</sup>Quaternions can only represent transformations in the SO(3) group, which contain all rotations, but no symmetry operations.

registration technique performed on points calculated via MDS techniques must be able to perform reflection transformations.

Registration calculates the rotation  $R$  and the translation  $\vec{t}$  that minimise the distance between the points  $x_i$  and  $y_i$ ,  $i \in 1 \dots n$ . In other words, it attempts to minimise:

$$\begin{aligned} d^2 &= \sum_{i=1}^n \|y_i - (Rx_i + \vec{t})\| \\ &= \sum_{i=1}^n \|y_i - Rx_i - \vec{t}\| \\ &= \sum_{i=1}^n (y_i - Rx_i - \vec{t})^T (y_i - Rx_i - \vec{t}) \end{aligned} \quad (12)$$

Translation, the easier of the two to calculate, is solved for first. Begin by calculating the centre of mass of both the  $x_i$  and  $y_i$ .

$$\begin{aligned} x_0 &= \frac{1}{n} \sum_{i=1}^n x_i \\ y_0 &= \frac{1}{n} \sum_{i=1}^n y_i \end{aligned}$$

While later it will be assumed that  $x_0$  and  $y_0$  are zero, we will begin by explicitly placing them in the equations in such a way that the rotations occur around the origin. This will be useful for solving for the translation vector,  $\vec{t}$ .

They are placed in the equation for  $d^2$  by substituting into equation (12), and then factorising:

$$\begin{aligned} d^2 &= \sum_{i=1}^n \left( (y_i - y_0) - R(x_i - x_0) + y_0 - Rx_0 - \vec{t} \right)^T \\ &\quad \left( (y_i - y_0) - R(x_i - x_0) + y_0 - Rx_0 - \vec{t} \right) \end{aligned} \quad (13)$$

Next, expand the product and remove unnecessary terms from the sum, to obtain

$$\begin{aligned} d^2 &= \sum_{i=1}^n \left( (y_i - y_0) - R(x_i - x_0) \right)^T \\ &\quad \left( (y_i - y_0) - R(x_i - x_0) \right) \\ &\quad + n(y_0 - Rx_0 - \vec{t})^T (y_0 - Rx_0 - \vec{t}) \end{aligned} \quad (14)$$

$\vec{t}$  has now been separated into a single term,  $n(y_0 - Rx_0 - \vec{t})(y_0 - Rx_0 - \vec{t})^T$ . Since the goal of the registration is to minimise  $d^2$ , the best way to achieve this in terms of  $\vec{t}$  is to set the term to

zero, in other words:

$$0 = n(y_0 - Rx_0 - \vec{t})^T(y_0 - Rx_0 - \vec{t}) \quad (15)$$

$$\Rightarrow \vec{t} = y_0 - Rx_0 \quad (16)$$

The  $R$  rotation matrix is determined next. First, define the  $n \times p$  matrices  $X$  and  $Y$ , which contain the  $n$  points  $x_i$  and  $y_i$  as their rows. Assuming that  $x_0$  and  $y_0$  are zero, and that  $\vec{t}$  has its value determined by equation 16; then from equation 14 we have

$$\begin{aligned} d^2 &= \sum_{i=1}^n (y_i - Rx_i)^T(y_i - Rx_i) \\ &= \sum_{i=1}^n y_i^T y_i + \sum_{i=1}^n x_i^T x_i - 2 \sum_{i=1}^n x_i^T R^T y_i \quad (\text{Since } R^T R = I) \\ &= \text{tr}(YY^T) + \text{tr}(XX^T) - 2\text{tr}(XR^TY^T) \end{aligned} \quad (17)$$

To minimise  $d^2$  we now have to maximise  $\text{tr}(XR^TY^T)$ , or, equivalently,  $\text{tr}(R^TY^TX)$  (remember that  $\text{tr}(AB) = \text{tr}(BA)$ ). Let  $C = Y^TX$ , and let the SVD of  $C$  be

$$C = U\Lambda V^T$$

with  $U$  and  $V$  being orthonormal matrices, and  $\Lambda$  the matrix of singular values. Substituting this into  $\text{tr}(R^TY^TX)$ , we obtain

$$\begin{aligned} \text{tr}(R^TY^TX) &= \text{tr}(R^TC) \\ &= \text{tr}(R^TU\Lambda V^T) \\ &= \text{tr}(V^TR^TU\Lambda) \end{aligned} \quad (18)$$

When multiplying  $\Lambda$  by an orthonormal matrix,  $\mathcal{X}$ , the product  $\mathcal{X}\Lambda$  must, along its diagonal, have elements that are less than or equal to the corresponding elements of  $\Lambda$ . This is clear, as the components of both  $\mathcal{X}$  and  $\Lambda$  are less than or equal to 1 (this is true for both matrices by definition:  $\Lambda$  being a collection of normalised eigenvalues, and  $\mathcal{X}$  being orthonormal). From this, and the fact that  $V$  (and hence  $V^T$ ),  $U$  and  $R^T$  are orthonormal, we can see that:

$$\text{tr}(R^TC) = \text{tr}(V^TR^TU\Lambda) \leq \text{tr}(\Lambda)$$

So  $\text{tr}(XR^TY^T)$  is maximised (i.e.,  $d^2$  is minimised) when  $\text{tr}(R^TC) = \text{tr}(\Lambda)$ . Then,

$$\begin{aligned} \text{tr}(R^TC) &= \text{tr}(\Lambda) \\ \Rightarrow \text{tr}(V^TR^TU\Lambda) &= \text{tr}(\Lambda) \\ \Rightarrow V^TR^TU\Lambda &= \Lambda \end{aligned} \quad (19)$$

Recall that an orthonormal matrix has a well defined inverse, which is its transpose; the inverse of a diagonal matrix  $\mathcal{X}$  is the diagonal matrix whose elements are the inverse of the corresponding elements in  $\mathcal{X}$ :

$$\begin{aligned} R^T &= VU^T \\ \Rightarrow R &= UV^T \end{aligned}$$

This rotation matrix will now maximise  $\text{tr}(XR^TY^T)$ , and hence minimise  $d^2$ . ■

And thus  $R$  and  $\vec{t}$  may be found. This rotation matrix and translation vector, when applied to the  $x_i$ s, aligns them with the corresponding  $y_i$ s. Because of the use of SVD in the calculation, if there is no exact translation and rotation to align the point sets, the technique calculates the rotation and translation that will bring the point sets as close together as possible, in a least squares sense.

This technique can be trivially extended to take scaling into account (Cox and Cox, 1994), as follows.

### Procrustes Superimposition

Registration itself only takes into account rigid transformations: in other words, those effected by rotations and translations. These changes effect only the *shape* of an individual's *form*. Size (scale) is not effected. *Procrustes superimposition* calculates both the rigid body transformation (in order to determine equivalence between shapes) and a scaling factor (to determine equivalence in size). Procrustes is essentially a registration technique that also takes into account isotropic scaling between the two point sets. The technique is typically used to align sets of landmarks from two individuals so that the differences between their landmark positions may be more readily examined. This is often the first step in studying the difference in form between two individuals when using landmark based morphometrics (Bookstein, 1991). Registration between organic forms is usually considered a difficult problem, with the details of any study strongly dependant on the landmarks chosen for registration (Bookstein, 1991; O'Higgins, 2000), although it is possible to determine gross differences in form (O'Higgins, 2000).

The SVD registration technique described previously has laid the foundation for developing a form of Procrustes based on least squared optimisation between the distances of the homologous landmarks. What remains is determination of scaling differences, which is a fairly trivial undertaking:

First, the optimisation function,  $d^2$ , must be recast to include scale:

$$d^2 = \sum_{i=1}^n \|y_i - (\sigma Rx_i + \vec{t})\|$$

$$= \sum_{i=1}^n (y_i - \sigma Rx_i - \vec{t})^T (y_i - \sigma Rx_i - \vec{t})$$

where  $\sigma$  is the new scaling factor. It is a simple exercise (left to the reader) to show that now

$$\vec{t} = y_0 - \sigma Rx_0$$

while  $R$  is unchanged.

Equation 17, incorporating scale, becomes:

$$d^2 = \text{tr}(YY^T) + \sigma^2 \text{tr}(XX^T) - 2\sigma \text{tr}(XR^TY^T)$$

Which, when differentiated with respect to  $\sigma$  is:

$$\frac{dd^2}{d\sigma} = 2\sigma \text{tr}(XX^T) - 2\text{tr}(XR^TY^T)$$

Setting  $\frac{dd^2}{d\sigma}$  to zero,

$$\sigma = \frac{\text{tr}(XR^TY^T)}{\text{tr}(XX^T)} \quad \blacksquare$$

### A Step by Step Approach to Registration and Procrustes Superimposition

The input to the technique are two sets of  $p$  dimensional points,  $\{x_i\}$  and  $\{y_i\}$ ,  $1 \leq i \leq n$ . The point  $x_i$  corresponds (i.e., is homologous to) the point  $y_i$ , for all  $i$ . Registration of  $\{x_i\}$  onto  $\{y_i\}$  proceeds as follows:

1. Calculate the centre of mass,  $x_0$  and  $y_0$ , for each point set.
2. Subtract the appropriate centre of mass from each point, forming  $x'_i = x_i - x_0$ , and  $y'_i = y_i - y_0$ .
3. Form the  $n \times p$  matrices,  $X$  and  $Y$ , such that row  $i$  of the matrix contains point  $x'_i$  or  $y'_i$ , respectively.
4. Calculate the singular value decomposition of the matrix  $Y^T X = U \Lambda V^T$ .
5. The rotation matrix is given by  $R = UV^T$ .
6. If required, calculate the scale factor  $\sigma$  using

$$\sigma = \frac{\text{tr}(XR^TY^T)}{\text{tr}(XX^T)}$$

7. Calculate the required translation using the equation  $\vec{t} = y_0 - \sigma Rx_0$ . If the scale factor  $\sigma$  is not required, and hence has not been calculated, set it to one.
8. Then apply the scale factor, rotation and translation to each  $x_i$ :

$$x''_i = \sigma Rx_i + \vec{t}$$

The  $x''_i$  values have been transformed to lie as “close” as possible to the corresponding  $y_i$ s in a least squares sense. As above, if the scale factor is not required, set it to one.

### 5.1.5 Centroid Size

Shape is defined to be the property of a form that is invariant under affine transformation, essentially translation, rotation and scale effects<sup>4</sup>. Size, however, remains an important consideration. The common measure of size for landmark data is the so-called *centroid size*. This is simple to compute: all the form’s landmarks are averaged to determine their centre of mass (i.e., their *centroid*). The sum of the squared distances from each landmark to the centroid is calculated, and the square root of this sum is taken. This value is the form’s centroid size.

#### A Step by Step Approach to Calculating Centroid Size

The input to this technique is a set of  $n$  landmarks,  $\mathcal{L} = \{l_i\}$ ,  $1 \leq i \leq n$ .

1. Calculate their centroid,

$$\bar{\mathcal{L}} = \frac{1}{n} \sum_{i=1}^n l_i$$

2. Their centroid size is given by,

$$\mathcal{L}_s = \sqrt{\sum_{i=1}^n |l_i - \bar{\mathcal{L}}|^2}$$

## 5.2 Reconstruction Techniques

The discussion now turns towards the three techniques which will be compared in the following chapters. These are:

---

<sup>4</sup>While translation, rotation and scale effects are usually all that are listed (for instance, see Dryden and Mardia, 1998), shape is also invariant under shear effects.

- *Mean substitution*, which concerns itself with replacing a missing landmark with its expected value, namely the landmark's average position calculated over a sample of undamaged reference specimens (section 5.2.1).
- *Thin plate spline correction*, which takes a reference specimen — possibly an average form<sup>5</sup> as with mean substitution — and calculates a homology map between the subset of landmarks common to both the reference and damaged specimens. While this homology map is calculated at these common landmarks, once determined it also specifies a mapping between all possible points — in other words, landmark and interlandmark points — on the damaged and reference individuals, including the extra landmarks on the reference specimen. These extra landmarks are homologous to the missing landmarks of the damaged individual, and so their transformation via the homology map is an estimate of the missing landmarks.
- *Correction via multiple linear regression*, a technique that calculates a set of regression models between the landmarks of a reference sample of individuals. These regression models capture the natural patterns of variation and covariation between the landmarks, and are used to estimate the position of a damaged individual's missing landmarks while taking into consideration not only the reference sample, but also the known portions of the damaged individual's morphology (section 5.2.3 on page 82).

### 5.2.1 Mean Substitution

Gunz (2005, pg 89, original emphasis) says in his thesis that,

It is appropriate right at the start to dismiss a method that is found in the literature: the *method of MEAN SUBSTITUTION* borrowed from the social sciences. ... This procedure makes no sense either as statistics or as science.

We will not argue over the technique's merits to either "statistics" or "science", but the results from Gunz (2005) suggest that mean substitution should perform poorly. In this way it stands as a benchmark: of the remaining two techniques, both are more costly to perform than mean substitution. If mean substitution performs as poorly as the work of Gunz (2005) implies, then in the cases where any other method is unable to outperform mean substitution one could ask why the extra effort should be spent.

The application of mean substitution is simple: first, a *reference form* is constructed. This is done using either a single individual<sup>6</sup> or a *consensus form*. The consensus form is a *Procrustes*

---

<sup>5</sup>Recall that a specimen's *form* is the collection of landmark measurements representing the specimen. See page 20.

<sup>6</sup>Unfortunately, due to the small samples sizes generally available to researchers, this is a frequent occurrence in fossil reconstruction. However, one must always bear in mind that a single data point cannot constitute a mean — in other

*average*: multiple individuals are aligned and scaled to each other using Procrustes superimposition; each group of homologous landmarks is then averaged, creating an average form. This initial Procrustes average is then iteratively improved: all the individuals are now aligned to the previously calculated average, and the mean of their homologous landmarks is once again calculated. This is repeated until the difference between the previous and current Procrustes average is sufficiently small.

The correction of the damaged specimen is trivial. The reference form is aligned with the damaged individual using common landmarks. The damaged form's missing landmarks are then estimated by substituting in the homologous landmarks from the reference form. An extra scaling step may also be employed: the reference form is scaled to match the damaged individual's centroid size, thus as far as possible removing scale effects. This scaling is performed throughout the work presented here.

Some general observations about the technique can be made: if the same reference individual / Procrustes average is used to correct multiple individuals, the various reconstructions may have identical areas of anatomy. Mean substitution shares a similarity to composite reconstructions. Composite reconstruction is a physical reconstruction technique that uses material from multiple individuals to produce one complete reconstruction. This is essentially mean substitution using a single reference individual. Mirroring techniques which replace missing anatomy by substituting the anatomy from the individual's symmetrical side is a special case of mean substitution.

### A Step by Step Approach to Performing Corrections Using Mean Substitution

The inputs are a damaged form and  $n$  reference forms. Steps 2 through 6 below calculate the consensus form. The correction is performed from step 7 onwards.

1. If there is only one reference form, proceed to step 7.
2. Choose a small value,  $\epsilon \in \mathbb{R}^+$ .
3. Select one of the reference forms as an initial approximation to the average.
4. For every form to be averaged,
  - (a) Calculate the Procrustes transformation of the form onto the average, as in section 5.1.4.
  - (b) Transform all the form's landmarks using this Procrustes transformation.

---

words, it is not a reliable measure of a landmark's expected value. This practice should always be avoided, and this is easily done when, for instance, in those cases where a modern human individual is used as the reference form.

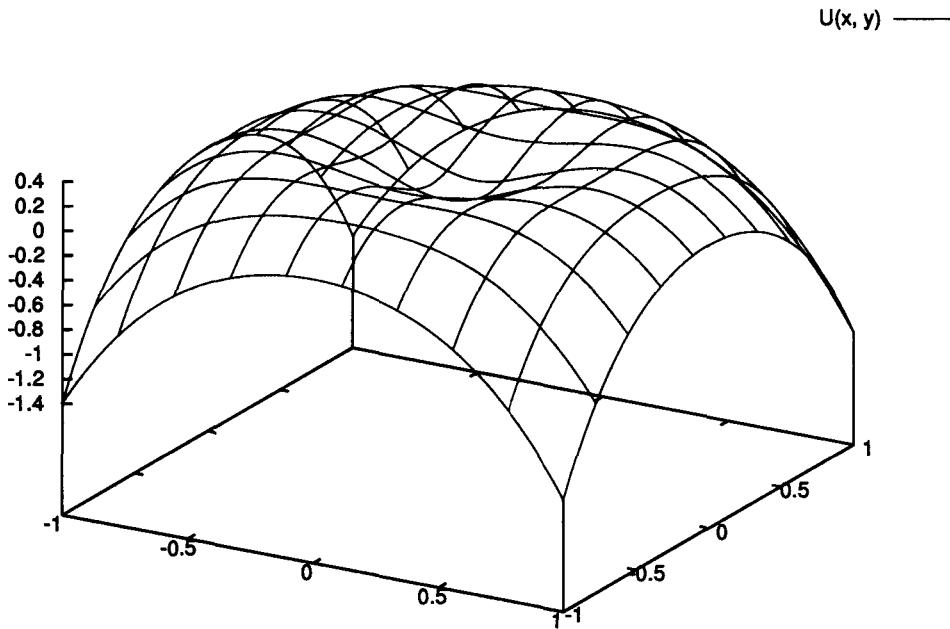


Figure 16: The function  $U(x, y) = -(x^2 + y^2) \log(x^2 + y^2)$

5. Calculate a new average form by grouping all the homologous landmarks and then finding each group's average.
6. Repeat from step 4 using the new average form. The process may be stopped when, for all the landmarks in the average form, the Euclidean distance between the landmark in its current and previous iteration is smaller than  $\epsilon$ . The current iteration is then the Procrustes average, and is used as the reference form.
7. Scale the landmarks of the reference form such that the form's centroid size is the same as that of the damaged form.
8. Align the reference form to the damaged form using common landmarks.
9. For each of the damaged form's missing landmarks, substitute the homologous landmark from the reference form.

### 5.2.2 Thin Plate Spline Warping

Thin plate splines (TPS) are a classic technique used for interpolating between two sets,  $R \subseteq \mathbb{R}^2$  and  $S \subseteq \mathbb{R}^2$ . Thin plate splines are functions of the form  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ . When used to interpolate

between points in  $\mathbb{R}^2$ , two such functions are used, one each for the  $x$  and  $y$  components. This pair of functions is defined on finite subsets of  $R$  and  $S$ ; the subsets are of equal size, and represent the corresponding (i.e., homologous) points of  $R$  and  $S$ . Once defined, the splines have as their domain the whole of  $R$ . They will exactly map any of the points from  $R$  initially used to define the function to its corresponding point in  $S$ . All the remaining points are mapped to  $S$  in such a way as to minimise a “bending energy” metric, as defined in equation 1 on page 22.

Thin plate splines originally found wide application in various fields outside of morphometrics (Richtsmeier et al., 1992), to which they were introduced by Bookstein (1989, 1991, 1986). They had by then already found application in computer graphics by, for instance, Terzopoulos (1983) — although in this case not as an interpolant. Their application to reconstruction is as an extension to mean substitution: a reference specimen or consensus form is taken to represent undamaged morphology. This form is then warped to fit the remaining undamaged portions of the damaged specimen. The warped reference form is then substituted to fill in the damaged areas. The warping via thin plate splines is assumed to reduce the error associated with standard mean substitution.

Thin plate splines obtained their name through their visual interpretation: the 2D case can be seen as a lofting of points from  $\mathbb{R}^2$  into a third dimension, creating a surface in  $\mathbb{R}^3$ . It is analogous to vertically displacing a thin sheet of metal at these points, and noting the surface defined by the deformation of this metal “plate”. If these displaced points lie almost in a plane, the shape of the plate minimises its “bending energy”, the same quantity that thin plate splines are designed to minimise (Small, 1996).

If there are  $n$  points in  $\mathbb{R}^2$ ,  $(x_i, y_i)$ , then a thin plate spline is a function of the form

$$f(x, y) = a_1 + a_2x + a_3y + \sum_{i=1}^n w_i U(x_i - x, y_i - y)) \quad (20)$$

$a_1, a_2, a_3$  and  $w_i$  (for  $1 \leq i \leq n$ ) being real valued.  $U$  is the function

$$U(x, y) = -(x^2 + y^2) \log(x^2 + y^2)$$

as shown in figure 16. The visualisation of  $f$  is the surface  $\{(x, y, f(x, y)) | (x, y) \in \mathbb{R}^2\}$  in  $\mathbb{R}^3$ .

With a modification to  $U$ ,  $f$  can be extended to  $\mathbb{R}^3$  in a straightforward manner:

$$f(x, y, z) = a_1 + a_2x + a_3y + a_4 + z \sum_{i=1}^n w_i U_{\mathbb{R}^3}(x_i - x, y_i - y, z_i - z)) \quad (21)$$

Here,  $U_{\mathbb{R}^3}(x, y, z) = \sqrt{(x^2 + y^2 + z^2)}$ , which is simply the point’s distance from the origin.

Since  $f$  is a function from  $\mathbb{R}^2$  into  $\mathbb{R}$ , two such functions are required when thin plate splines are used as an interpolant, one for the  $x$  and  $y$  coordinates, respectively. Naturally, three functions are required in the  $\mathbb{R}^3$  case. These will be labelled as  $f_x$ ,  $f_y$  and  $f_z$ .

A thin plate spline is completely defined by the real valued constants  $a_1, a_2, a_3, a_4$  and  $w_i$  for  $1 \leq i \leq n$ , which may be calculated using linear algebra, as below.

Let  $n \in \mathbb{N}^+$ , and  $x_i = (x_{ix}, x_{iy}, x_{iz})$ , for  $1 \leq i \leq n$ , be landmarks on the *reference* form, and  $y_i = (y_{ix}, y_{iy}, y_{iz})$  the homologous landmarks on the *target* form. Construct the following matrices:

$$X = \begin{pmatrix} 1 & x_{1x} & x_{1y} & z_{1z} \\ 1 & x_{2x} & x_{2y} & z_{2z} \\ 1 & x_{3x} & x_{3y} & z_{3z} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{nx} & x_{ny} & z_{nz} \end{pmatrix} \quad (22)$$

and

$$Y = \begin{pmatrix} y_{1x} & y_{1y} & z_{1z} \\ y_{2x} & y_{2y} & z_{2z} \\ y_{3x} & y_{3y} & z_{3z} \\ \vdots & \vdots & \vdots \\ y_{nx} & y_{ny} & z_{nz} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (23)$$

Let  $\mathcal{D} = (d_{ij})$  be the  $n \times n$  matrix whose diagonal is all zero, and whose remaining elements are  $d_{ij} = U_{\mathbb{R}^3}(x_{ix} - x_{jx}, x_{iy} - x_{jy}, x_{iz} - x_{jz})$ <sup>7</sup>. Then compose the matrix

$$L = \left( \begin{array}{c|c} \mathcal{D} & X \\ \hline X^T & 0 \end{array} \right) \quad (24)$$

Finally, calculate the product:

$$L^{-1}Y = \begin{pmatrix} w_{1x} & w_{1y} & w_{1z} \\ w_{2x} & w_{2y} & w_{2z} \\ \vdots & \vdots & \vdots \\ w_{nx} & w_{ny} & w_{nz} \\ a_{1x} & a_{1y} & a_{1z} \\ a_{2x} & a_{2y} & a_{2z} \\ a_{3x} & a_{3y} & a_{3z} \\ a_{4x} & a_{4y} & a_{4z} \end{pmatrix}$$

---

<sup>7</sup>In  $\mathbb{R}^3$ , this is simply the form matrix; see chapter 2.

The constants for each of the three TPS functions are the columns of the matrix  $L^{-1}Y$ .

An important property to note is that each landmark  $x_i$  is mapped onto its corresponding  $y_i$ ; in other words,  $f_x(x_{ix}, x_{iy}, x_{iz}) = y_{ix}$ , and similarly for  $f_y$  and  $f_z$ . This ensures that homologous landmarks remain homologous.

Reconstruction of a specimen — whose landmarks make up the points of the *target* form — is now posed as an interpolation problem to be solved using thin plate splines. First, a *reference* form is chosen. As with mean substitution, this may either be a single specimen or a consensus form. Landmarks common to both forms are used to define a pair or a triplet of TPS functions,  $(f_x, f_y, f_z)$ . The damaged form's missing landmarks are then estimated by transforming their homologous landmarks on the reference form via the thin plate spline.

### A Step by Step Approach to Performing Corrections Using Thin Plate Splines

As input, we have one damaged form and  $n$  reference forms.

1. If there is only one reference form, proceed to step 3.
2. Calculate a consensus form from the  $n$  references as with steps 2 through 6 on page 77. This consensus form is now used as the sole reference form for the remainder of the reconstruction.
3. Align the reference form to the damaged form.
4. Using only the landmarks common to both the damaged and reference forms, create matrices  $X$  and  $Y$ , as in equations 22 and 23.
5. Calculate  $L$  as in equation 24.
6. Calculate  $L^{-1}Y$ . The columns of this matrix define the coefficients for the needed thin plate splines.
7. Use the above coefficients to construct three thin plate splines (two in the 2D case),  $f_x, f_y$  and  $f_z$ .
8. For each of the damaged form's missing landmarks,
  - (a) Locate the homologous landmark,  $l$ , in the reference form.
  - (b) Apply  $f_x, f_y$  and  $f_z$  to  $l$ , obtaining the estimate for the missing landmark.

### 5.2.3 Multiple Linear Regression

The main disadvantage of a technique such as mean substitution is that it only consults a reference sample in order to model how the missing portions of a specimen should appear. It makes no use of the information contained in any of the damaged specimen's undamaged areas, and in doing so ignores any constraints that these areas may impose on the missing portions. While thin plate spline methods make a limited attempt to use undamaged morphology (by essentially forcing another individual's shape to fit the undamaged portions of the individual under correction), regression methods go a step farther by using the relationships between landmarks to predict the position of missing landmarks. To do this, regression methods build a model of the variation / covariation inherent between the landmarks of a reference sample. The relationship between the damaged specimen's known landmarks can then be used in conjunction with this model to predict the position of missing landmarks.

The use of regression methods poses a number of challenges. The most immediate concerns sample sizes: the calculation of the regression coefficients requires a reference sample size of at least the same size as the set of predictor variables. In the case presented here, the more landmarks we wish to use as predictors — hence allowing the model to capture and be contingent on more of the morphology's variation / covariation — the larger the required sample of undamaged reference specimens. Considering the small number of fossil specimens, and the even smaller subset of *undamaged* fossil material, this is clearly problematic.

The approach used here is based on the *coordinate-free* method of Richtsmeier et al. (1992). Rather than performing regression analyses on position coordinates, this technique operates on their respective matrix of interlandmark distances, the so called *form matrix* introduced in chapter 2. Distances are a quantification of the relationship between the landmark positions; their use removes the need to align the reference individuals into a common space. In the form matrix, landmarks are related by a row and column of distance entries to the other landmarks. Missing landmarks correspond to a missing row and column. A subset of the remaining distances are then used as predictor variables in a regression estimation of the missing distance variables.

$$\begin{array}{c}
 \text{To Landmark} \\
 \hline
 \text{From Landmark} \left\{ \begin{pmatrix} 0 & d_{12} & \cdots & d_{1n-1} & d_{1n} \\ d_{21} & 0 & \cdots & d_{2n-1} & d_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{n-1,n} & 0 \end{pmatrix} \right\}
 \end{array}$$

However, the number of predictor variables dictates the required size of the reference sample. When using a coordinate-based regression technique (i.e., using the landmarks directly in the regression analysis) on  $n$  landmarks, the number of predictor variables required to estimate a single missing landmark is  $3 \cdot (n - 1)$ , there being  $n - 1$  non-missing landmarks, each with three position components (i.e.,  $x$ ,  $y$  and  $z$ ) to a 3D landmark. The number of variables are linear in  $n$ , the number of landmarks. The form matrix, on the other hand, contains a distance variable for each pair of landmarks, creating an  $n \times n$  matrix of distance variables. Even with the estimated landmark's row and column missing, the number of variables is still quadratic in  $n$ . Clearly, a coordinate-free approach produces more predictor variables — and consequently a requirement for larger reference samples — than does a coordinate-based approach.

However, there are certain properties of the form matrix that can be exploited:

1. The distance variables on the diagonal are always zero, being the distance from a landmark to itself. This information is meaningless to the regression analysis.
2. The matrix is symmetric. In other words, for a given form matrix  $\mathcal{D} = (d_{ij})$ , we know that  $d_{ij} = d_{ji}$  for all  $i$  and  $j$ .
3. Richtsmeier et al. (1992) point out that the distance variables are highly interdependent. This is self evident: consider a triplet of landmark coordinates and their interlandmark distances. The relationship between the distances is clear if one considers the three landmarks as defining a triangle, the interlandmark distances being the respective lengths of the triangle's sides. Then the distances are related by the standard trigonometric relationships, and do not satisfy the need for independence of predictor variables required by standard multiple linear regression methods.

Points 1 and 2 allow for a simple reduction in the number of variables. The effect of point 3 is less clear. Multiple linear regression methods require that the predictor variables be independent of each other. A dependency between the predictor variables often allows for variables to be removed without reducing the predictive ability of the regression model. Indeed, removing these variables can improve the regression model as a whole (e.g., Farrar and Glauber, 1967; Jolliffe, 1972; Leahy, 2000).

One possible method for reducing the number of variables is through the use of a principal component analysis (PCA) on the distance variables; this supplies a sense of which variables explain the most variance in the data set, and hence which can more readily be removed without reducing the regression model's predictive abilities (Jolliffe, 1972, 1973, 1986). The technique has a twofold value. Firstly, the existence of eigenvalues of zero value in a PCA analysis is an indicator that some

variables are multicollinear (Jolliffe, 1986) — these are variables whose values are completely determined by the value of other predictor variables. Distance variables associated with these eigenvalues may be completely removed, as they add nothing to the regression model. Secondly, after the removal of all multicollinear variables, variables associated with the smaller eigenvalues may also be removed, until a suitable number of variables remain relative to the size of the available reference sample. It should be noted that the removal of variables associated with points 1 and 2 above, along with the removal of multicollinearities, is usually sufficient to reduce the number of variables to be equivalent to coordinate-based methods<sup>8</sup>.

One important question to be asked is why make use of a coordinate-free approach when it requires a larger reference sample? The important feature of a coordinate-free approach is the *resolution* at which the damaged and undamaged morphology is modelled. When the reconstruction is performed directly on the landmarks, the landmarks are either considered completely present or completely missing (e.g., Gunz, 2005). However, this does not capture the reality of damaged morphology. Some landmarks may be grouped together because they are in their correct positions relative to the other landmarks within the group, but when considering the landmarks contained in two such groups, damaged morphology lying between the groups will mean that the landmarks are no longer in their correct position relative to those of the other group. The most extreme example of this is the fragmentation of friable fossil material, a poor choice for reconstruction via techniques such as mean substitution and thin plate spline methods as presented here, but easily handled by a coordinate-free regression method<sup>9</sup>. In other words, landmarks may not be completely present or missing, but lie somewhere in between.

This stems from the fact that the landmarks themselves are merely points. Landmarks cannot be “damaged” or “undamaged”, although they may lie on damaged regions of the fossil material (in which case all of their relationships to their surrounding landmarks are incorrect, and the landmark may be considered to have no information content and be completely missing), or the regions that lie between two landmarks may be damaged (and thus only the relationship between landmarks that cross this damaged area is lost, and all other relationships remain intact — the landmark still retains some information, and is, in a sense, neither completely present nor missing). Coordinate-based methods must either completely remove semi-present landmarks, or they must subdivide the fossil specimen into portions without ambiguity as to whether a landmark is missing or not.

---

<sup>8</sup>Unfortunately a proof of why this should be so is not available. However, it seems that exactly the same information concerning shape (i.e., everything invariant under translation, rotation and scale operations) is captured by both landmark coordinates and their associated distance matrix. This being so, a possible conjecture is that if one of these representations produces more variables for analyses than another then some of this information is redundant, since the same information can be captured by the smaller number of variables.

<sup>9</sup>A coordinate-free regression method will preserve the interlandmark distances on the various fragments, while estimating those between fragments. This will then position the fragments relative to one another. An exercise for the reader is the application of a coordinate-free mean substitution method to such a problem.

Neither of these two approaches, however, is satisfactory, since neither allows a reconstruction to make full use of a specimen's known, undamaged morphology. The ability of a coordinate-free regression method to treat only the distances effected by taphonomic distortion as missing, rather than whole landmarks, is an important, practical advantage of the technique over coordinate-based methods. The difficulty related to increased sample sizes and the practicalities of variable reduction seem a mild trade off for the possible improvements in the reconstruction process and output. One may be concerned that the variable reduction removes the advantages obtained by using coordinate-free methods, but the variable reduction is not random and unguided. It is based on some knowledge of which predictor variables offer either no information (the multicollinear variables), or very little (those associated with small eigenvalues). The PCA method is also only one of many variable reduction approaches, any of which may be used at the researcher's prerogative<sup>10</sup>.

The PCA method used here is straightforward, and taken from Jolliffe (1972, 1973, 1986). The principal components of all the distance variables are calculated, supplying a set of eigenvectors and their associated eigenvalues. The eigenvalues indicate the importance of their eigenvectors: those associated with smaller eigenvalues explain less of the variance contained within the data than eigenvectors with larger eigenvalues. A cutoff value  $\lambda \in \mathbb{R}^+$  is chosen. All eigenvalues of smaller value are considered to explain little enough variance that we may wish to remove the distance variables associated with these eigenvalues. Determining the associated distance variable is achieved through the eigenvectors. If there are  $n$  landmarks, then the eigenvectors will be of dimension  $n$ . It is known that the  $i^{th}$  component of an eigenvector is associated with the  $i^{th}$  landmark. The method presented here associates a distance variable to the whole eigenvector by locating the eigenvector's largest component, then choosing that component's associated distance variable. If this distance variable has previously been removed by association to another eigenvector, the eigenvector's second largest component is found, its distance variable removed, and so on, until all the eigenvalues below  $\lambda$  have had a distance variable removed.

The work here uses a cutoff value of  $\lambda = 0.0005$ , much smaller than  $\lambda = 0.7$  proposed by Jolliffe (1972). This value reduced the number of distance variables to the same number of variables produced by the coordinate-based regression method.

These variables are then used as the predictors in a series of regression models, one model for each missing distance variable.

---

<sup>10</sup>For example, principal component regression techniques, or techniques that examine the correlation between the dependant and independent variables.

### A Step by Step Approach to Performing Corrections Using Multiple Linear Regression

Given a set of  $n$  reference individuals, each represented as a set of  $j$  landmarks,  $\{l_i | l_i \in \mathbb{R}^3 \text{ for } 1 \leq i \leq j\}$ . Also given is a single individual,  $X$ , missing either landmarks or distances. Steps 4 through 7 below perform the PCA variable reduction technique. The reconstruction is performed in steps 9 through 11.

1. Convert the landmarks of each individual into a form matrix  $\mathcal{D} = (d_{ij})$ , each  $d_{ij}$  being the distance between landmarks  $l_i$  and  $l_j$ . Missing landmarks are represented by the appropriately missing rows and columns of distance variables in the form matrix.
2. From each reference individual's form matrix, remove from consideration the upper triangle of elements (i.e., all  $d_{ij}$  where  $i < j$ ).
3. From each reference individual's form matrix, remove from consideration the diagonal elements,  $d_{ii}$ .
4. For each of the reference specimens, relabel the distance variables as  $d_i$  (i.e., drop a subscript). We can now represent each individual as a vector  $\vec{d} = (d_m)$ . This will result in  $n$  such vectors.
5. Calculate the principal components of these  $n$  vectors. This will result in a set of eigenvectors,  $\vec{e}_o$ , with a corresponding set of eigenvalues,  $e_o$ .
6. Select a suitably low cutoff point  $\lambda \in \mathbb{R}^+$ .
7. For each eigenvalue,  $e_o$ 
  - (a) If  $e_o < \lambda$ , consider the corresponding eigenvector  $\vec{e}_o$ . Locate the largest component of the eigenvector,  $\eta_k$ .
  - (b) The corresponding distance variable is  $d_k$ , which should be removed from consideration as a predictor. If  $d_k$  was already removed from consideration, repeat step 7a above, searching for the next largest component.
8. The remaining subset of distance variables are now used as predictor variables. For each missing variable in the damaged form  $X$ , calculate a standard multiple linear regression model on the reference forms using these predictor variables. This work made use of the `lm` regression function found in *R* (R Development Core Team, 2005).
9. For each missing variable in the damaged form matrix of  $X$ , estimate its value using the appropriate regression model.

10. Transform the now completed form matrix into landmarks using multidimensional scaling.  
Call this calculated set of landmarks  $X'$  (section 5.1.3).
11. Align  $X'$  to  $X$  using a registration method (section 5.1.4).

#### 5.2.4 Comments on the Methods

It is worth making some general observations about the methods as they stand. All of them make use of a reference sample. The information concerning each landmark in the sample is used:

- either individually, on a landmark by landmark basis, such as with mean substitution and thin plate splines, where individual landmarks are averaged.
- Or simultaneously, where the interlandmark relationships are considered, such as the variance / covariance patterns exploited by regression-based techniques.

The thin plate spline and regression methods take into consideration more than just an external data set specified by a reference sample: they also consider the known morphology of the damaged individual, either in order to fit the reference model (TPS), or for driving predictions based on variation / covariation (regression). Mean substitution may be augmented to make limited use of the known morphology of the damaged individual by considering the individual's size. This has been done in the work presented here, as explained in the section 5.2.1. Intuitively, one would expect those methods making the most use of the damaged individual's known morphology to achieve the best reconstructions, although chapter 6 shows that this is not always the case.

#### 5.2.5 Correcting Interlandmark Morphology

The standard correction of interlandmark morphology is performed using thin plate splines. As with landmark corrections, a homology map is defined between the landmarks shared by the reference and damaged individuals. Interlandmark morphology is typically measured using Computed Tomography (the work of Glenn Conroy is an example: Conroy and Vannier, 1987, 1989; Conroy et al., 1998, 2000), and thin plate splines provide a mapping from the undamaged morphology of the reference specimen to that of the damaged individual. This technique has been used in a number of instances, for example by Zollikofer and Ponce de León (2005); Zollikofer et al. (2002a), where a badly fractured cranium, also suffering from plastic distortion, was reconstructed; thin plate splines were used to remove the plastic distortion from the cranium by creating an homology map between the landmarks of the original, distorted cranium, and its correction. In their reconstruction of the Le Moustier 1 Neandertal specimen, Ponce De León and Zollikofer (1999) completed the

specimen's endocranial surface by fitting the homologous surface from a modern human, and, in a comparative attempt during the same analysis, anatomy from the Gibraltar 1 Neandertal specimen. The interlandmark morphology was required to estimate the individual's endocranial capacity.

This estimation is typically performed using thin plate splines by determining a homology maps, as suggested by Bookstein (1991). The use of thin plate splines is fairly arbitrary. Thin plate splines do have useful qualities (previously mentioned), however they lack biological validity, since two forms have no biological imperative that the homology map between them should maintain a minimum bending energy. Indeed, it is possible for thin plate splines to create a homology map that display *foldover effects*, which cause anatomy to intersect with itself. However, TPS are easy to compute, and do not have constants that need to be modified by a user in order to obtain decent results.

All of the quantitative analyses in the remaining chapters test only the three landmark reconstruction techniques just presented. The following chapter, chapter 6, compares the performance of these techniques in a series of corrections on members of extant ape species. Chapter 7 discusses the results and makes recommendations on which technique to use in a given circumstance. The recommendations are concerned with estimating missing landmarks, and should be used to do so. Thin plate splines can then be employed to reconstruct missing interlandmark morphology.

# Chapter 6

## Tests and Results

Before employing any reconstruction technique, it is important to have a sense of both its behaviour and utility. This chapter studies the behaviour of the three techniques employed in this work, examining their relationship to the number of missing landmarks to be estimated, the number of individuals in the reference sample, and the use of living species as an alternate — and larger — reference sample. The results to the various analyses are reported here<sup>1</sup> and in appendix A; they are discussed in the following chapter, in which the methods are compared, and where we make recommendations concerning the use of these techniques. The results are there also examined to see how they answer the research questions given in chapter 1.

### 6.1 Materials and Methods

Because of the scarcity of fossil material, the bulk of the testing is performed using data from individuals of three extant primate species: *Homo sapiens* ( $n = 10$ ), *Pan troglodytes* ( $n = 10$ ) and *Gorilla gorilla* ( $n = 10$ )<sup>2</sup>. These samples are the *test samples* on which the corrections are performed. The correction techniques themselves also require a *reference sample* to guide the corrections (see chapter 5 for details); the number of individuals used for the reference sample varies across analyses (since the reference sample size is often an independent variable under examination), however each analysis states this number. There is a maximum of 57 chimp, 97 gorilla, and 600 human individuals available from which to draw the reference samples. All individuals are adult, and all the samples have an equal number of male and female individuals — or approximate, in the case of an odd

---

<sup>1</sup>Statistical analyses and graphing were performed with R (R Development Core Team, 2005).

<sup>2</sup>Collected in a previous study (Ackermann, 1998).

Landmark	Description	Position
NA	Nasion	midline
NSL	Nasale	midline
ANS	Anterior nasal spine	midline
IS	Intradentale superior	midline
FMN	Frontal-maxillary-nasal junction	left, right
ZS	Zygomatico-maxillare superior	left, right
ZI	Zygomatico-maxillare inferior	left, right
FM	Fronto-malare	left, right
ZTS	Zygo-temporal superior	left, right
ZTI	Zygo-temporal inferior	left, right
MT	Maxillary tuberosity	left, right
PT	Pterion	left, right
TSP	Temporo-sphenoidal-parietal junction	left, right
AS	Asterion	left, right
PNS	Posterior nasal spine	midline
BA	Basion	midline
OPI	Opisthion	midline
BR	Bregma	midline
LD	Lambda	midline

Table 2: The landmarks recorded for each individual. Of these, most of the analyses are performed using only the facial landmarks: NA, NSL, IS, FMN, ZI, FM, ZTS, MT.

number of individuals<sup>3</sup>. The test and reference samples do not overlap. Each individual used in an analysis is represented by up to 29 landmark measurements, shown in table 2. Of these, most of the analyses are carried out on the subset of thirteen facial landmarks — NA, NSL, IS, FMN, ZI, FM, ZTS, MT — in order to reduce the number of variables involved in regression-based reconstructions (see section 7.2.3). All landmark measurements were obtained through averaged repeated measures using a Microscribe contact digitiser.

The goal of the testing is to compare and contrast the various correction methods. This is done by performing the same correction with all three methods on each member of a test sample. Residuals between the true and corrected landmarks are then calculated. The mean of each sample's residuals are compared for statistically significant differences. Welch's approximate *t*-test for heteroscedastic data is employed as the significance test of choice, since an F-test shows that the variances of the different residual samples are unequal (Sokal and Rohlf, 1995).

---

<sup>3</sup>Individuals of each sex were pooled for two reasons: first, not enough members of one sex alone were available for the analyses, and secondly, researchers may not be sure of the sex of a fossil individual, making the use of pooled samples appropriate.

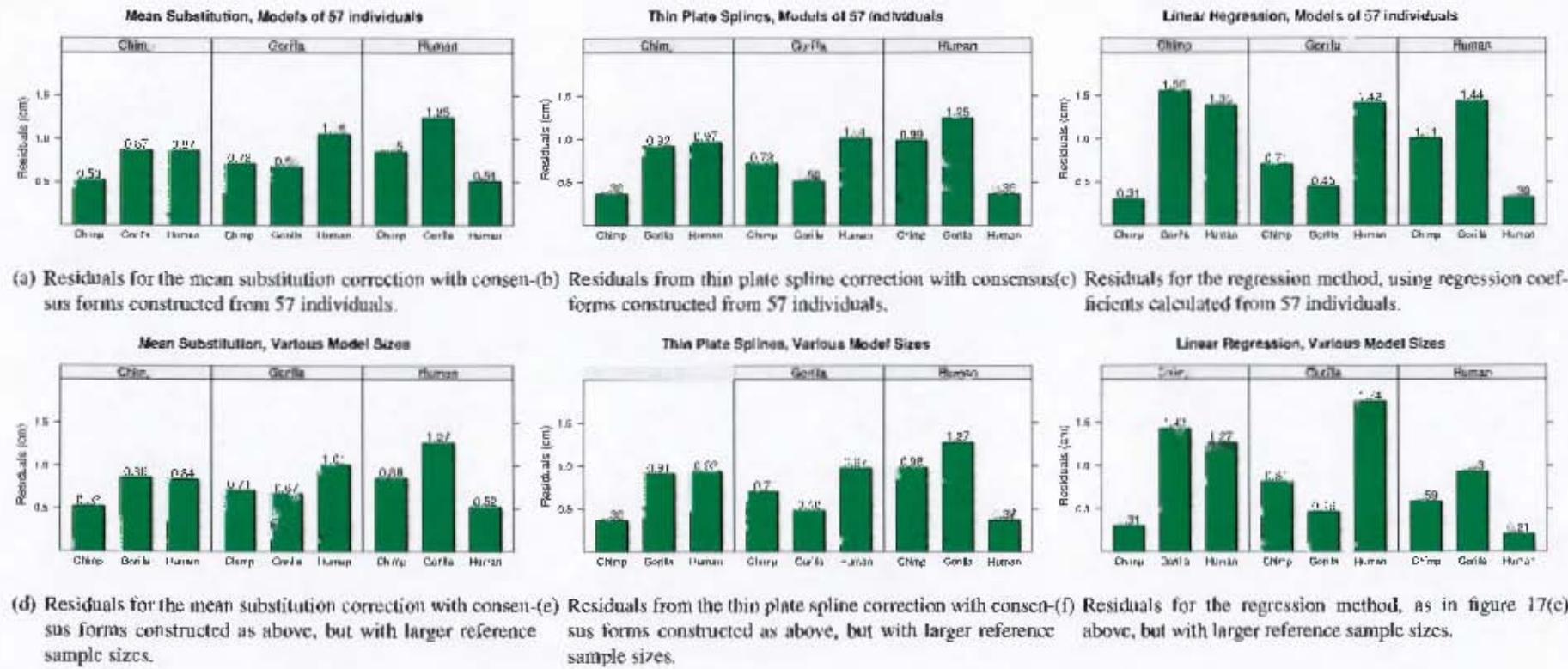


Figure 17: These graphs show the residuals obtained for the mean substitution (section 6.1.1), thin plate spline (section 6.1.2) and regression-based (section 6.1.3) techniques. Corrections are performed using within- and across-species reference samples. The x-axis shows the species of the individual being corrected; the groupings give the species from which the reference sample is drawn. The y-axis shows the mean residuals. The bottom row of graphs are calculated as for top row, but using larger reference sample sizes.

- (b) Calculate the individual's mean residual.
3. Calculate the mean residual across all individuals.

Note that the mean calculated here is a mean of individuals:

$$\mu = \frac{1}{57} \sum_{i=1}^{57} \left( \frac{1}{13} \sum_{j=1}^{13} \rho_{ij} \right) \quad (25)$$

where  $\rho_{ij}$  is the residual for  $i^{th}$  individual and the  $j^{th}$  landmark. The inner summation calculates the mean for an individual, while the outer summation calculates the mean across individuals. Using this information it is easy to calculate a mean of *landmarks* — in other words, given a particular landmark, calculate its average error. This is done by noting that one can expand the summations and multiply the constants through, then refactorise to produce the mean calculated over landmarks:

$$\mu = \frac{1}{13} \sum_{j=1}^{13} \left( \frac{1}{57} \sum_{i=1}^{57} r_{ij} \right) \quad (26)$$

The inner summation is the mean residual for a given landmark. The outer summation is the mean across all the landmarks. It is important to note that the above equations are equivalent — they calculate the same value,  $\mu$ , and contain the same information. All the means for most of the analyses have been calculated using equation 25 — however, the reader should note the equivalence.

Figure 17(a) shows the obtained mean residuals. The overall mean for the method, across all test and reference sample combinations, is  $\bar{X} = 0.815\text{mm}$ . For comparison with later regression tests (section 6.1.3), means obtained from larger reference sample sizes are supplied in figure 17(d) (chimp reference sample size = 57, gorilla = 97, human = 280). With these reference samples, the method obtains a total mean of  $\bar{X} = 0.812\text{mm}$ . There is no significant difference between these two means (observed  $P = 0.930$ ), and further *t*-tests between the means obtained for each test sample while varying the reference sample size again shows no significant difference (observed  $P$  values all above 0.3). Table 10 on page 137, as well as tables 11, 14 and 15, show the difference in the means between mean substitution and the other two methods (for which results are obtained in analyses 6.1.2 and 6.1.3).

### 6.1.2 Analysis II — The Thin Plate Splines Method

This analysis reconstructs the test samples using the thin plate spline warping correction method (described in section 5.2.2). Again, three reference samples of gorilla, human and chimp individuals are used, consisting of 57 individuals each.

The test proceeds as follows:

1. Create a consensus form by aligning and averaging the reference individuals' landmarks.
2. For each test individual,
  - (a) For each landmark,
    - i. Remove the landmark.
    - ii. Align the consensus form to the remaining landmarks.
    - iii. Create thin plate splines from the consensus form's landmarks onto the individual's remaining landmarks.
    - iv. Approximate the individual's missing landmark by warping the corresponding landmark of the consensus form with the splines.
    - v. Calculate the residual between the true and the corrected landmark.
  - (b) Calculate the individual's mean residual.
3. Calculate the mean residual across all individuals.

Figure 17(b) lists the means. The method's total mean, across all test and reference sample combinations, is  $\bar{X} = 0.789\text{mm}$ . As with mean substitution, we repeat the spline corrections with larger reference samples (chimp reference size = 57, gorilla reference size = 97, human reference size = 280). The total mean in this case is  $\bar{X} = 0.779\text{mm}$ , and the results are displayed in figure 17(e). As with mean substitution, there is no significant difference between these means (observed  $P = 0.804$ ), nor between the individual test samples as the reference samples are changed ( $P$  values all above 0.3). Table 12 on page 138, and tables 13, 14 and 15, show  $t$ -tests for differences in the mean between the results of analyses I and II, and analyses II and III.

### 6.1.3 Analysis III — The Regression Method

This analysis corrects the test samples using the regression-based technique described in section 5.2.3. Three reference samples of 57 individuals — one each of gorilla, chimp and human individuals — are used.

The analysis proceeds as follows:

1. Calculate a set of regression coefficients.
2. For each individual in the test sample,
  - (a) For each landmark,
    - i. Remove the landmark.

- ii. Transform the landmark configuration into a distance matrix with the row and column of the removed landmark similarly missing.
- iii. Use the regression coefficients to estimate the missing entries in the distance matrix.
- iv. Calculate a set of points corresponding to the distances (via multidimensional scaling — see section 5.1.3).
- v. Rotate, translate and mirror these newly calculated points so that they overlay their original, corresponding landmarks.
- vi. Substitute the extra landmark for the missing landmark.
- vii. Calculate a residual.

(b) Calculate a mean residual across all of the landmarks.

3. Calculate a mean residual across all the individuals.

Mean residuals are given in figure 17(c), with the total mean for the technique being  $\bar{X} = 0.957\text{mm}$ . The analysis is repeated using larger reference samples (chimp = 57, gorilla = 97 and human = 280), with results in figure 17(f), and obtaining a total mean of  $\bar{X} = 0.988\text{mm}$ . The observed  $P$ -value between these means is 0.109, which is almost significant at the 0.1 level, but is not significant at the 0.05 level used here. The difference between the means obtained for test samples corrected with the smaller and larger human reference samples are all significantly different (observed  $P < 0.03$ ), as is the human test sample corrected with the gorilla sample. All other differences are not significant (observed  $P > 0.3$ ).  $t$ -test comparisons to the mean substitution and thin plate spline results are provided in table 11 on page 138, as well as tables 10, 12, 13.

#### 6.1.4 Analysis IV — Cross Over Point

We wish to know how small a reference sample the regression-based technique requires for it to become competitive with the mean substitution and thin plate spline warping techniques. To discover this, the test sample is repeatedly corrected while we increase the reference sample sizes in increments of ten individuals, starting at ten individuals and ending at 600. Only a human reference sample is used for this analysis, since it is the only data set of large enough size available to us. There are once again human, chimp and gorilla test samples ( $n = 28$ ; since no individuals were needed to construct reference samples for the chimp and gorilla data sets, some of the individuals could be used in the test samples). The corrections were performed as outlined in analyses I through III. Figure 18 displays the results obtained from this analysis, overlaying the mean substitution, spline and regression residuals. Table 16 on page 141 shows  $t$ -test comparisons between the regression and mean substitution methods; table 17 compares the regression and spline residuals, and table 18 compares the spline and mean substitution results.

### Comparison With Increasing Reference Sample Sizes

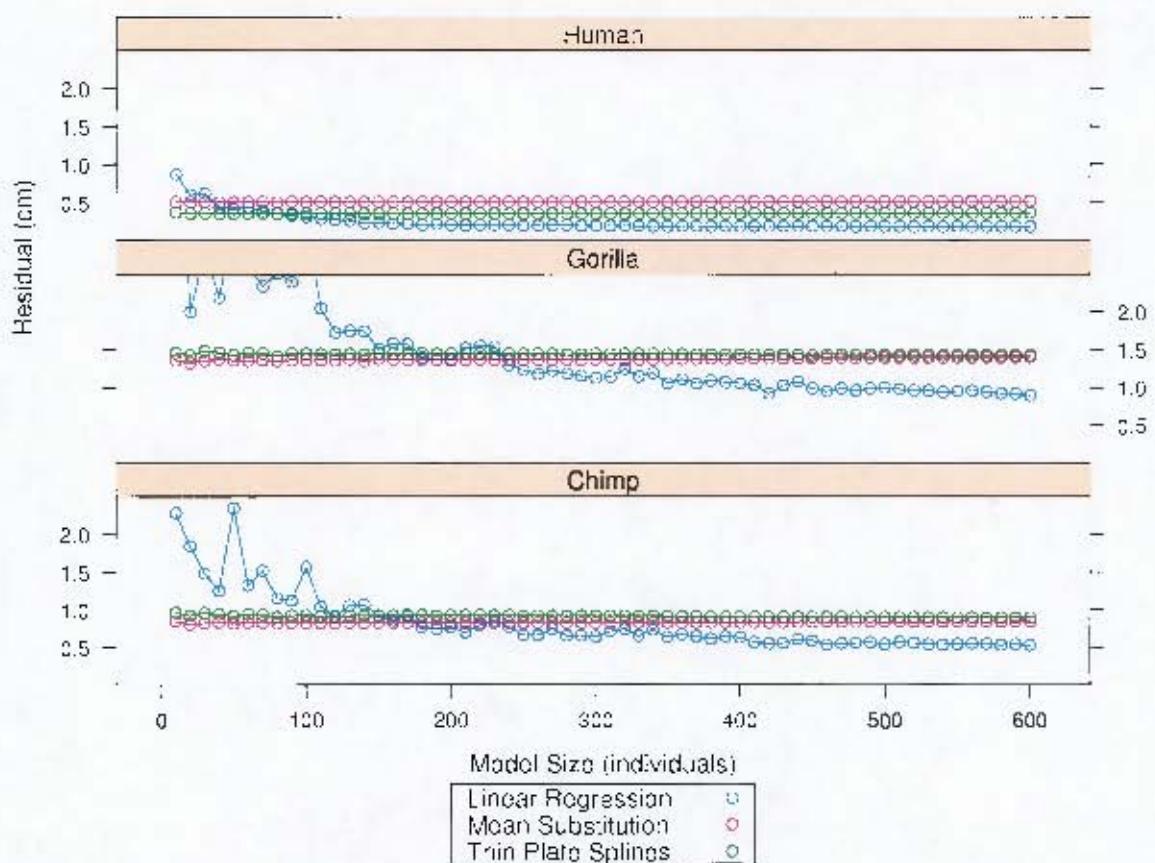


Figure 18: The graph shows a comparison between the mean substitution, thin plate spline and regression techniques. All corrections were performed with a human reference sample on human, chimp and gorilla test samples. Each plot point is the sample average of each individual's mean landmark residual (the  $y$ -axis). The  $x$ -axis shows the reference sample sizes. The reference sample sizes increase in increments of ten, from ten up to 600 individuals. Note that mean substitution and thin plate splines do not show a reduction in residual sizes as the number of individuals used to create the consensus form increases: for large reference samples, their corrective power appears invariant to the sample size (this comes with the same proviso regarding outliers that all techniques associated with a mean must come with). The regression-based method, however, clearly shows a reduction in residual size.

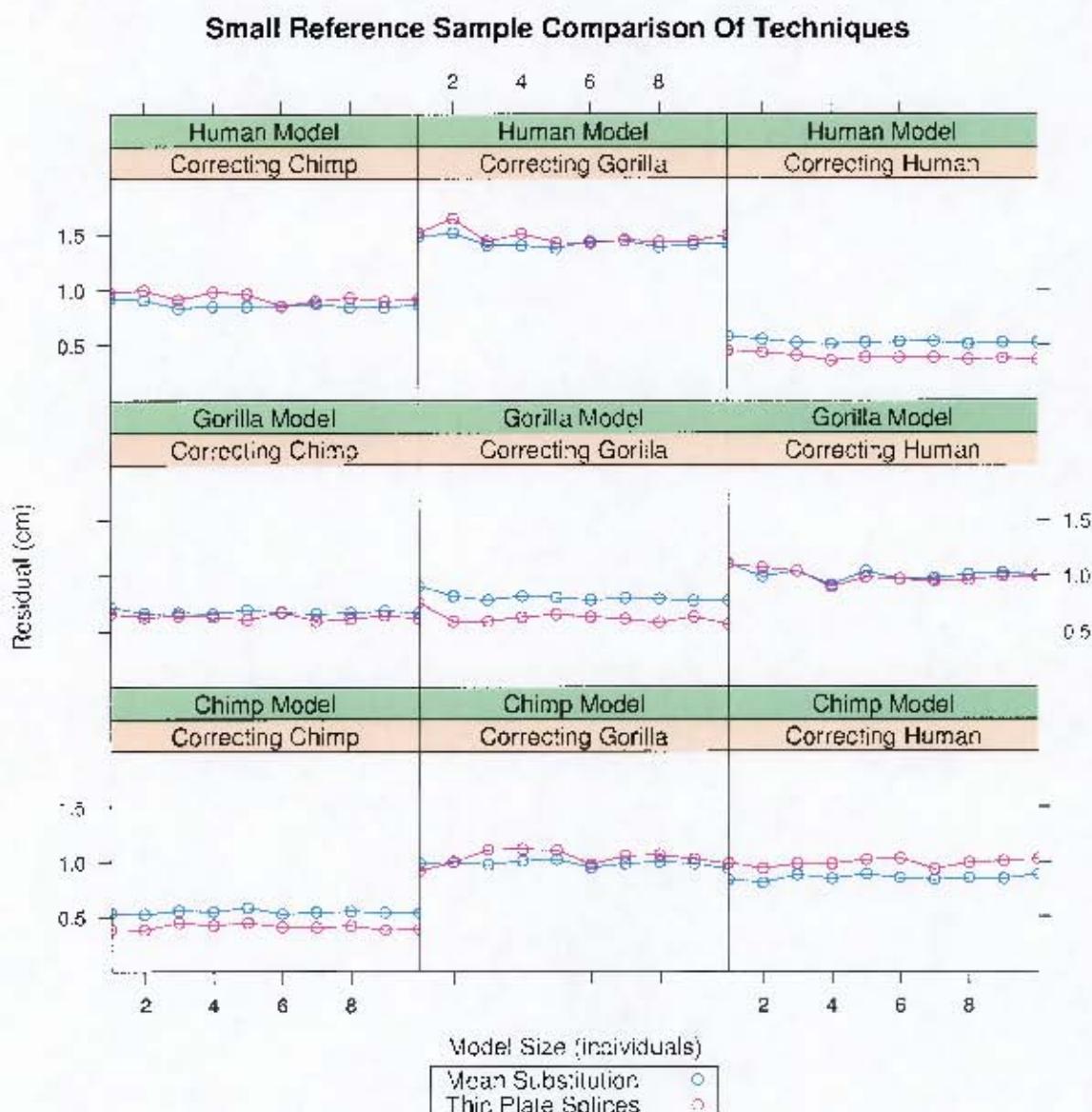


Figure 19: The graph is similar to that of figure 18; it compares the mean substitution and thin plate spline methods as the reference sample sizes increase from one individual up to ten (figure 18 compares the techniques for larger reference samples). The regression-based method is not included in this graph, since with such a small reference sample its predictive power is extremely poor for the number of variables involved in the reconstruction. The *x*-axis plots the reference sample size, the *y*-axis the mean residual. Each test sample has  $n = 28$ .

Figure 19 shows the test repeated using small reference samples; from one up to ten individuals, in single increments. This test allows us to see how mean substitution and thin plate spline methods react to small reference sample sizes (since both rely on the construction of a mean, which is sensitive to outliers, especially when sample sizes are small; also, small reference sample sizes are unfortunately the norm, rather than the exception). The regression method requires a much larger reference sample size to be competitive and is not included in the analysis. With the smaller reference samples, enough data is available for us to test the two techniques with reference samples drawn from each of the three species. Tables 19, 20 and 21 give significance tests between the two techniques' means for the chimp, gorilla and human reference samples respectively. Table 24 compares the means obtained using the regression-based technique and the human reference sample of 600 individuals against the means obtained using mean substitution and thin plate splines and the smaller chimp and gorilla reference samples. This is done in order to see if a large reference sample can make across-species corrections competitive to within-species corrections using techniques based on small reference samples.

### 6.1.5 Analysis V — Cumulative Errors

Analyses I through III calculate landmark residuals as if only a single landmark were missing. Unfortunately, taphonomic distortion typically effects multiple landmarks on a single specimen; this test examines how the three techniques react to this by iteratively increasing the number of landmarks requiring simultaneous correction. The number of missing landmarks can be thought of as a proxy for the amount of damage that a specimen suffers from.

All the test individuals ( $n = 28$ ) are corrected using a human reference sample of 600 individuals, allowing the regression-based method to perform without the hindrance of small sample sizes. The test is repeated ten times for each individual, with each iteration successively removing between one and ten landmarks. The test proceeds as follows:

1. Create a consensus form from the reference sample.
2. Let us say that  $m$  is the number of landmarks being corrected, and let  $m$  initially have the value one.
3. For each test sample of chimp, gorilla and human individuals ( $n = 28$ ),
  - (a) For each individual,
    - i. Remove  $m$  random landmarks from each individual.
    - ii. Create a family of regression coefficients that do not make use of the  $m$  removed landmarks.

### Cumulative Error Growth in the Correction Techniques

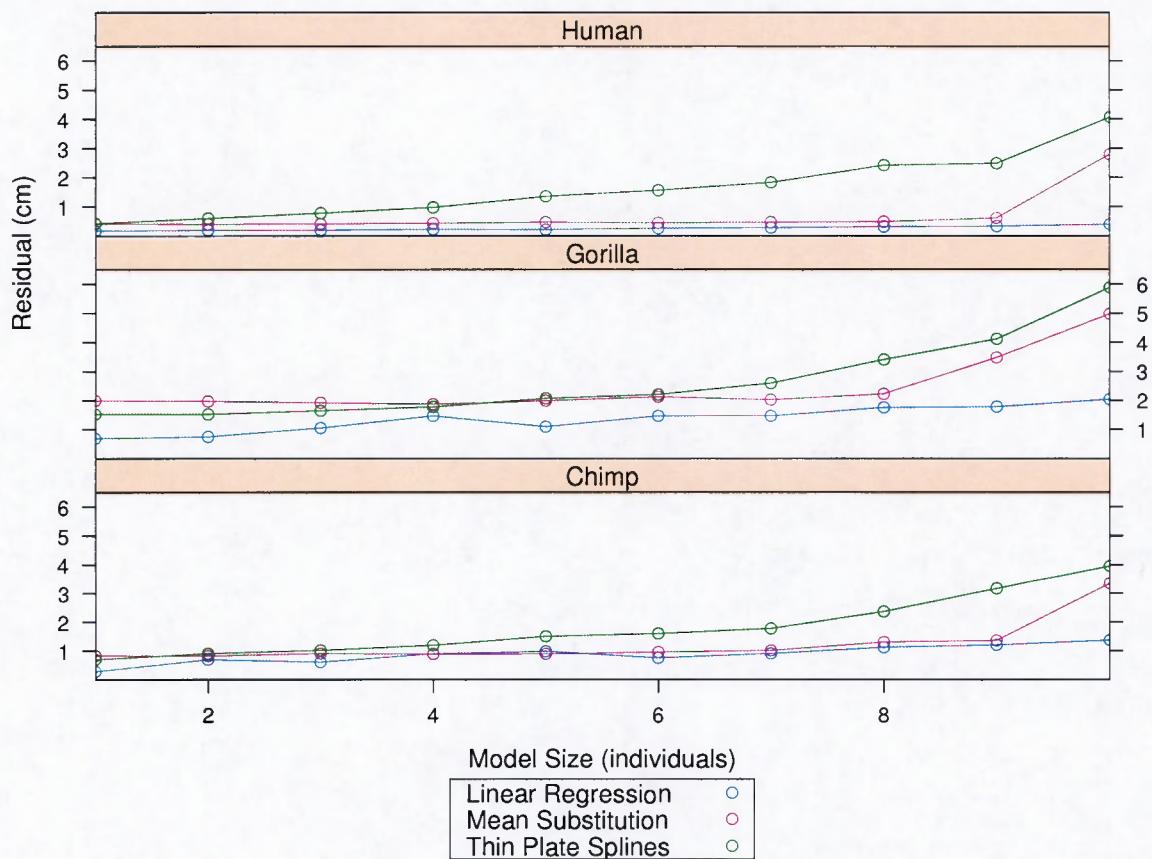


Figure 20: This displays a comparison between the three correction techniques as landmarks are cumulatively removed from the test individuals. The reference sample consists of 600 modern human individuals. Tests are conducted by removing from one to ten landmarks from the thirteen used in the analysis (shown on the  $x$ -axis). The  $y$ -axis displays the calculated mean residual.

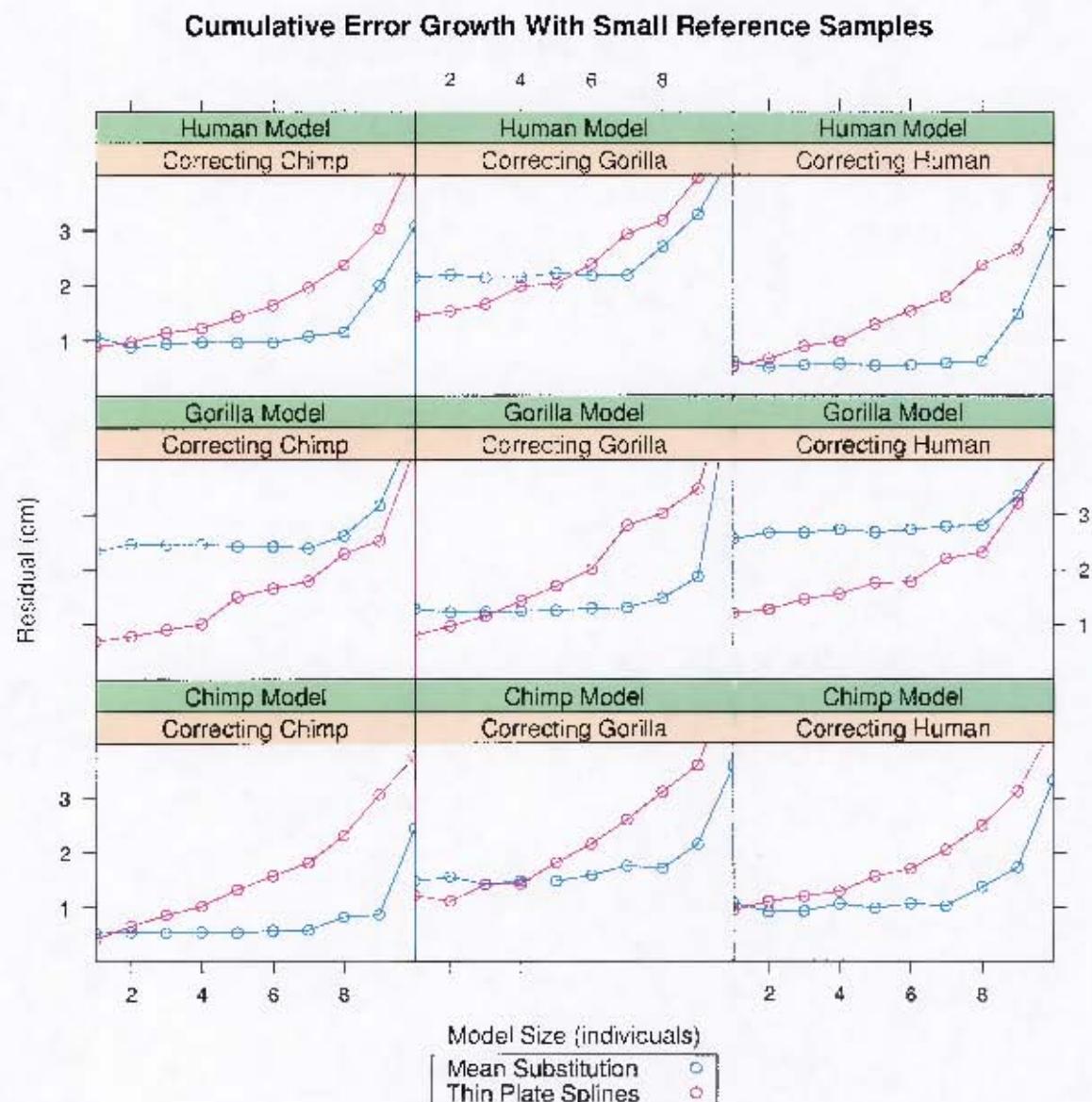


Figure 21: This displays a comparison between mean substitution and thin plate splines while landmarks are cumulatively removed from the test individuals. It is similar to figure 20, except that the reference sample consists of only one randomly selected individual. The regression method is excluded, as the reference sample of size one is not sufficient to make use of the technique. Test samples have  $n = 28$ . The  $x$ -axis plots the number of landmarks being corrected, and the  $y$ -axis the mean residual.

- iii. Correct each of the  $m$  landmarks simultaneously with a correction method.
  - iv. Calculate the residual for each of the corrected landmarks.
- (b) Calculate the mean residual.
4. Calculate the mean residual over all individuals. This is now the mean residual for correcting  $m$  landmarks.
  5. Increase the number of landmarks corrected ( $m$ ) by one, and repeat the process from step 3 onwards. Stop when  $m$  is larger than ten.

Only ten landmarks are removed, as some of the techniques require registration with the original, damaged, landmark configuration. Further, thin plate splines and regression equations require the test individual's undamaged landmarks as input (for use as predictors, or to define the thin plate splines). The ten landmark loss represents damage to 77% of the landmarks.

Figure 20 on page 99 shows the results of the analysis. Table 25 give the means obtained for each technique at each iteration of the analysis, while tables 27, 28 and 29 give *t*-test results comparing these means to one another.

It is interesting to consider when the mean obtained for correcting only a single landmark becomes significantly different from the later means (i.e., when does an increase in the number of damaged landmarks degrade the performance of the corrective technique). With a 0.05 *P*-value, this occurs at the following points:

- Mean substitution:
  - Chimp test data: **9** landmarks and onwards, although correcting **7** landmarks is also significantly different.
  - Gorilla test data: **9** landmarks and onwards.
  - Human test data: **10** landmarks.
- Thin plate spline warps:
  - Chimp test data: **3** landmarks and onwards.
  - Gorilla test data: **5** landmarks and onwards.
  - Human test data: **2** landmarks and onwards.
- Multiple linear regression:
  - Chimp test data: **2** landmarks and onwards.
  - Gorilla test data: **3** landmarks and onwards.

- Human test data: **6** landmarks and onwards.

We repeat the test using only a single reference individual, this being an often experienced case in hominin cranial reconstructions. The test is run as previously outlined, except for the following changes:

- The reference sample consists of one randomly selected individual.
- Because of the small reference samples, the regression-based technique is not used.
- The use of only a single reference individual allows us to perform each test three times, once each with a chimp, gorilla and human reference individual.

Figure 21 displays the results, and table 26 the associated means. Tables 30, 31 and 32 give the *t*-test results between the mean substitution and thin plate spline techniques for the various reference samples used in the analysis.

Again, we look for the number of missing landmarks at which the mean residual is significantly different ( $P < 0.05$ ) from the mean obtained for correcting only one landmark:

- Mean substitution:
  - Chimp reference sample:
    - \* Chimp test data: **9** landmarks and onwards.
    - \* Gorilla test data: **9** landmarks and onwards.
    - \* Human test data: **9** landmarks and onwards.
  - Gorilla reference sample:
    - \* Chimp test data: **9** landmarks and onwards.
    - \* Gorilla test data: **10** landmarks.
    - \* Human test data: **9** landmarks and onwards.
  - Human reference sample:
    - \* Chimp test data: **9** landmarks and onwards.
    - \* Gorilla test data: **9** landmarks and onwards.
    - \* Human test data: **9** landmarks and onwards.
- Thin plate spline warps:
  - Chimp reference sample:
    - \* Chimp test data: **2** landmarks and onwards.

- \* Gorilla test data: **5** landmarks and onwards.
- \* Human test data: **3** landmarks and onwards.
- Gorilla reference sample:
  - \* Chimp test data: **3** landmarks and onwards.
  - \* Gorilla test data: **3** landmarks and onwards.
  - \* Human test data: **3** landmarks and onwards.
- Human reference sample:
  - \* Chimp test data: **3** landmarks and onwards.
  - \* Gorilla test data: **4** landmarks and onwards.
  - \* Human test data: **2** landmarks and onwards.

It is also interesting to ask if there is a significant difference when using the larger reference sample. Table 33 lists the *t*-test results between the means obtained for the mean substitution technique using the large reference sample (figure 20) against the single reference individual (figure 21). Table 34 does the same for the thin plate spline method.

#### 6.1.6 Analysis VI — Landmark Error Spread

This is the one analysis in which we use the full complement of 29 landmarks. It allows us to see the error distribution over the whole cranium.

A human reference sample ( $n = 128$ ) is used to correct a human test sample ( $n = 33$ ). The analysis is carried out as follows:

1. Calculate a consensus form and a set of regression coefficients.
  - (a) For each individual in the test sample.
    - i. Remove a random landmark.
    - ii. Correct the landmark using a correction method (as in analyses I, II and III).
    - iii. Calculate a residual between the true and corrected landmark.
  - (b) Calculate the individual's mean residual.
2. Calculate the mean residual over all the individuals in the sample.

Results are given for mean substitution (figure 22), thin plate splines (figure 23) and multiple linear regression (figure 24) on the same test sample. Figures 22(a), 24(a) and 23(a) show the mean residuals per individual. Figures 22(b), 23(b) and 24(b) show the mean residuals per landmark. The

latter two figures also partition the landmarks into facial and non-facial landmarks, to aid the reader in examining the results.

The mean residual for the mean substitution method is  $\bar{X} = 0.47\text{mm}$ , for the regression method  $\bar{X} = 0.35\text{mm}$ , and for thin plate splines  $\bar{X} = 0.41\text{mm}$ . As previously noted, the mean is the same whether we calculate it across landmarks or individuals. Showing the residuals per landmark, though, allows us to partition the residuals between those acquired from facial and neuro / basicranial landmarks. The mean residual of the facial landmarks obtained via mean substitution is  $\bar{X}_F = 0.41\text{mm}$ , while the mean for the non facial landmarks is  $\bar{X}_{NF} = 0.51\text{mm}$ . One can ask if there is a significant difference between these means, and there is: a one sided *t*-test gives an observed  $P = 0.017$ ,  $df = 26.999$  and  $t = -2.544$ . Similarly, for the regression method  $\bar{X}_F = 0.29\text{mm}$  and  $\bar{X}_{NF} = 0.39\text{mm}$ . Again there is a significant difference between the mean: a one sided *t*-test supplies an observed  $P = 0.005$ ,  $df = 26.224$  and  $t = -3.094$ . The same also holds true for the spline warping technique:  $\bar{X}_F = 0.31\text{mm}$ ,  $\bar{X}_{NF} = 0.49\text{mm}$ . There is a significant difference between the means, with  $P = 0.001$ ,  $t = -4.107$ ,  $df = 20.306$ .

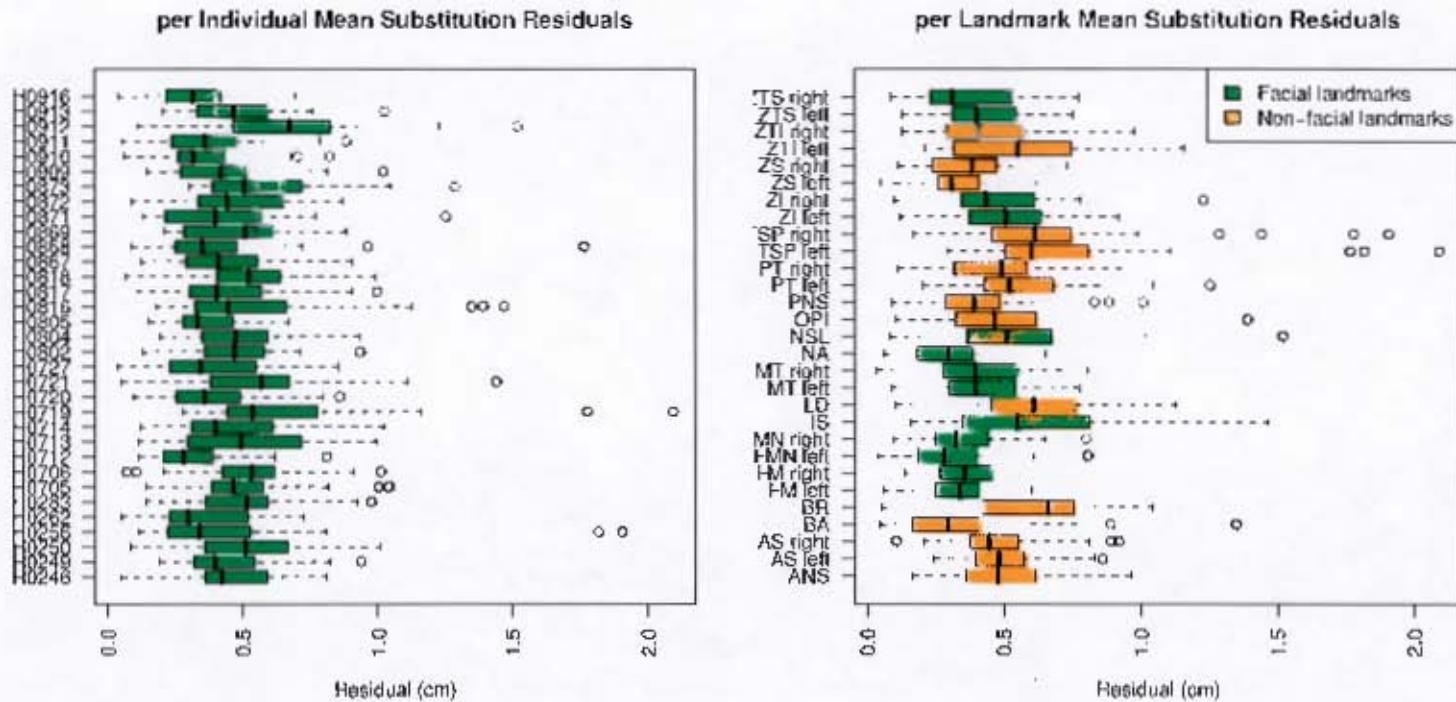
### 6.1.7 Analysis VII — Correcting Fossil Specimens

The previous analyses are all performed on test samples of living species. Most reconstructions, however, occur on fossil material from extinct species. This analysis attempts to examine the behaviour of the techniques when “correcting” fossil specimens. It is important to note that in this test we are not attempting to estimate landmarks that are truly missing from these fossils. Rather, we are attempting to estimate the known landmarks — in this way we are able to estimate the amount of reconstruction error associated with each technique as has been done in the previous analyses. The corrections are performed in the same manner as that of the extant species in analyses *I* through *III*, by removing known landmarks, in turn, and estimating them.

Individuals from various species are used in this analysis. The australopiths are represented by *Australopithecus africanus* (STS 5, Taung), *Paranthropus boisei* (KNM-ER 406, KNM-WT 17400), and *Paranthropus aethiopicus* (KNM-WT 17000). Our genus, *Homo*, is represented by *Homo habilis* (KNM-ER 1470, KNM-ER 1813) and *Homo erectus* (KNM-ER 3733, KNM-WT 15000). The specimens are listed in tables 4 and 5. Each specimen is in a different state of preservation, with some showing more or less damage than others<sup>4</sup>. This has effected which landmarks have been collected for which specimen; the tables list the landmarks collected for each individual.

Each specimen is corrected using all of the methods, as well as three reference samples: human ( $n = 628$ ), chimp ( $n = 67$ ) and gorilla ( $n = 117$ ). The corrections proceed as in analyses *I* through

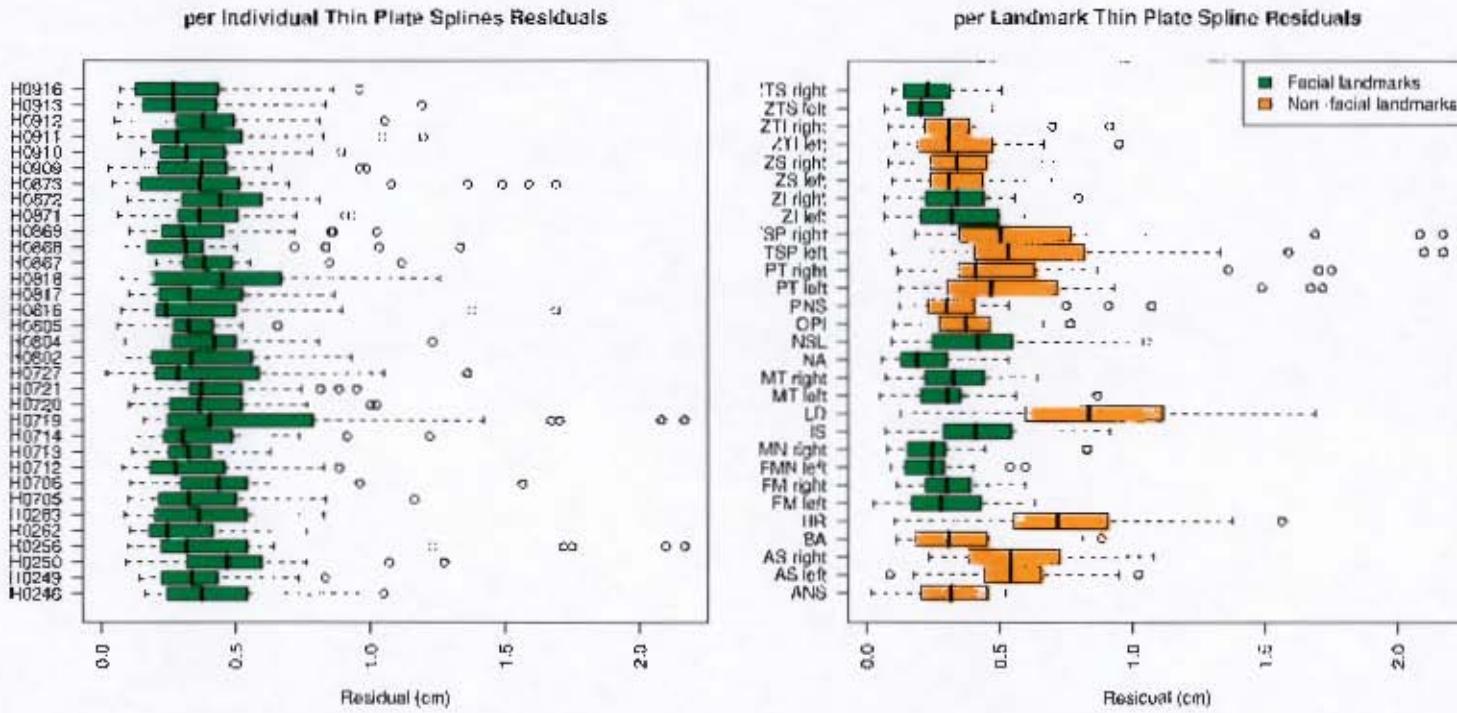
<sup>4</sup>While we are only estimating known landmarks in this analysis, it is perfectly possible to use the various techniques to estimate those landmarks which are truly missing. Given the results presented in this section, a human reference sample driving a regression-based reconstruction should provide a good degree of accuracy in estimating these landmarks.



(a) Box and whisker plots of the residuals obtained by using mean substitution to predict missing landmarks on test individuals. The y-axis shows the individuals. Each plot gives the residual range obtained for a particular individual across its landmark corrections. The total mean residual is  $\bar{X} = 0.47$ .

(b) A box and whisker plot showing the range of residuals obtained for each landmark, measured across individuals. The mean residual is identical to the total mean of the individuals:  $\bar{X} = 0.47$ . We can see that that the mean for the facial landmarks is lower than the mean for the non-facial landmarks. These are  $\bar{X}_F = 0.41$ , and  $\bar{X}_{NF} = 0.51$  respectively. Welch's approximate *t*-test show that these means are significantly different (observed  $P = 0.017$ ).

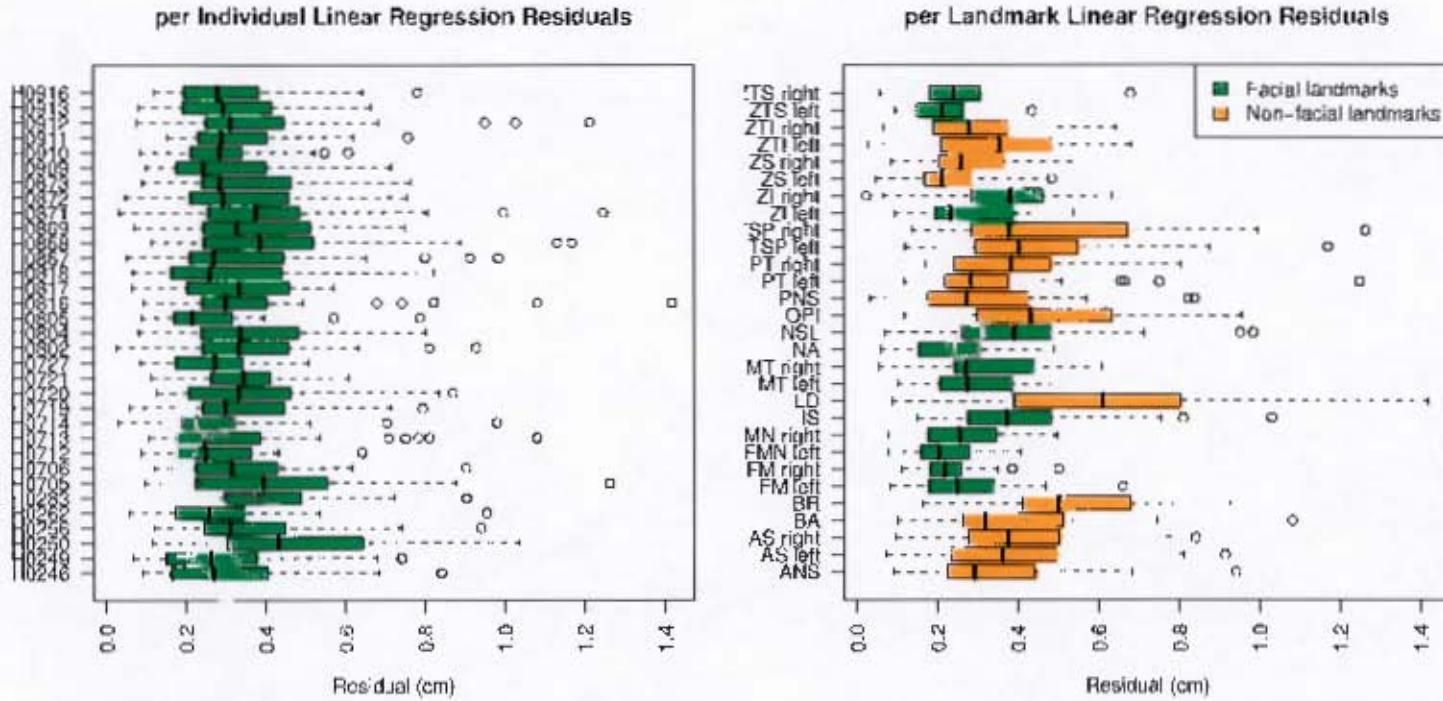
Figure 22: Error distributions for the mean substitution method.



(a) Box and whisker plots of the residuals obtained by using thin plate splines to predict missing landmarks. The y-axis shows the individuals. Each plot shows the range of residuals obtained for the particular individual as each of its landmarks is sequentially corrected. The mean residual is  $\bar{X} = 0.41\text{mm}$ .

(b) A box and whisker plot showing the range of residuals obtained for each landmark corrected using thin plate splines, measured across individuals. The total mean residual is identical to that calculated over the individuals:  $\bar{X} = 0.41\text{mm}$ . The facial landmarks' mean is lower than the mean for the non-facial landmarks:  $\bar{X}_F = 0.31\text{mm}$ , and  $\bar{X}_{NF} = 0.49\text{mm}$ . Welch's approximate *t*-test show that these means are significantly different (observed  $P = 0.001$ )

Figure 23: Error Distribution for the Thin Plate Spline Method.



(a) Box and whisker plots of the residuals obtained by using regression equations to predict missing landmarks on an individual. The y-axis shows the individuals. Each plot shows the range of residual obtained for the particular individual as each of its landmarks are sequentially estimated by the technique. The mean residual is  $\bar{X} = 0.35\text{mm}$ .

(b) A box and whisker plot showing the range of residuals obtained for each landmark corrected using regression equations, measured over individuals. The total mean of the residuals is identical to the total mean over all the individuals:  $\bar{X} = 0.35\text{mm}$ . As with the other techniques, the mean for the facial landmarks is lower than the mean for the non-facial landmarks:  $\bar{X}_F = 0.29\text{mm}$ , and  $\bar{X}_{NF} = 0.39\text{mm}$ . Welch's approximate *t*-test show that these means are significantly different (observed  $P = 0.005$ ).

Figure 24: Error distribution for the regression method.

*III*, each landmark being in turn removed then estimated. A mean residual is calculated for the technique. These means are reported in the tables.

This analysis has a number of problems: first, the varying number and selection of the landmarks may have some effect (varying the number of landmarks certainly effects how much information is available to techniques such as our regression-based method). Second, we cannot be completely sure if any slight plastic deformation of the specimens may be effecting — and by how much — any of the landmarks (more than likely there is some), and hence the calculated residuals. The fossils themselves are from various species on which the techniques may have more or less of a reconstructive ability.

With these reservations noted, we present the results in tables 4 and 5. Averaged over all three reference samples, mean substitution produces a total weighted mean of  $1.150\text{cm}$ , thin plate splines a mean of  $1.294\text{cm}$ , and multiple linear regression  $1.221\text{cm}$ . There are no significant differences between these means ( $P > 0.05$ ). However, the regression-based method, when using the human reference sample of  $n = 628$  individuals, produces a mean residual of only  $0.809\text{cm}$ , far below the other techniques ( $P < 0.05$ ).

The australopith specimens are morphologically closer to chimp and gorilla than to human; *Homo* is similar to human. Table 3 supplies mean residuals for the reconstructions calculated for genus, with the chimp / gorilla means combined into one group. We can test for differences in the means to see if the australopith individuals are better corrected by chimp / gorilla reference samples, and similarly if *Homo* is better corrected by a human reference sample. For the australopith group, the mean substitution and thin plate splines techniques show no difference between the residuals obtained for the chimp / gorilla reference group and the human reference sample ( $P > 0.05$ ). The regression technique produces a significantly smaller mean when using the human reference sample (which is the larger sample).

For the *Homo* group, mean substitution shows no difference ( $P > 0.05$ ) between the chimp / gorilla reference sample group, and the human reference sample. Both the thin plate spline and regression-based methods produce smaller residuals using the human reference sample. The mean residual for correcting *Homo* individuals via regression equations, using a human reference sample, is only  $0.666\text{ cm}$ .

The correction of the fossil specimens are also performed using single fossil specimens as reference individuals. In this case, only mean substitution and thin plate splines are employed. Table 6 presents corrections performed with STS 5 as the reference individual, table 7 uses KNM-ER 406, and table 8 KNM-ER 3733. For each reference individual there is no significant difference between the means obtained through thin plate splines and mean substitution ( $P > 0.05$ ).

Genus	Mean Subst			TPS			Regression		
	C	G	H	C	G	H	C	G	H
Australopith	1.121	1.237	1.263	1.328	1.322	1.321	1.984	1.323	0.909
	1.179	1.263		1.325		1.321	1.654		0.909
<i>Homo</i>	1.030	1.174	1.002	1.399	1.485	0.868	1.255	0.947	0.666
	1.102	1.002		1.442		0.868	1.101		0.666

Table 3: This table displays grouped weighted mean residuals for correcting the fossil specimens in tables 4 and 5. The columns represent the reconstruction methods and the species of the reference samples (C: chimp, G: gorilla, H: human). The rows give the genera the specimens have been grouped in. The “Australopith” group is made up of the *Paranthropus* and *Australopithecus* genera. Combined weighted mean residuals for the chimp and gorilla reference samples are also given.

The regression-based method, using the large human reference sample, produces a smaller mean residual than either mean substitution or thin plate splines when these methods employ single fossil specimens as a reference sample ( $P < 0.05$ ). Further, there is no significant difference in using mean substitution with either a living or extinct reference sample, as with thin plate splines ( $P > 0.05$ ).

Specimen	Species	No. of Lndmrks	Recorded Landmarks	Mean Subst			TPS			Regression		
				C	G	H	C	G	H	C	G	H
STS 5	<i>Australopithecus africanus</i>	12	NA, NSL, IS, FMN left, FMN right, ZI left, ZI right, FM left, FM right, ZTS right, MT left, MT right	0.842	0.931	1.088	0.824	0.920	1.192	1.357	0.691	0.768
KNM-ER 1470	<i>Homo habilis</i>	7	NA, NSL, IS, FMN left, FMN right, FM left, FM right	1.466	1.547	1.400	2.076	2.525	1.764	0.810	1.049	0.858
KNM-ER 1813	<i>Homo habilis</i>	8	NA, NSL, IS, FMN left, FMN right, FM right, MT left, MT right	0.916	1.167	0.945	1.119	1.409	1.032	1.080	0.889	0.529
KNM-ER 3733	<i>Homo erectus</i>	12	NA, NSL, IS, FMN left, FMN right, ZI left, ZI right, FM left, ZTS left, ZTS right, MT left, MT right	0.884	1.034	0.864	1.011	0.939	0.444	1.898	0.824	0.532
KNM-ER 406	<i>Paranthropus boisei</i>	13	NA, NSL, IS, FMN left, FMN right, ZI left, ZI right, FM left, FM right, ZTS left, ZTS right, MT left, MT right	1.616	1.695	1.638	1.960	1.601	1.570	3.446	2.992	1.462

Table 4: This table displays fossil specimens whose landmarks are estimated using mean substitution, thin plate splines, and multiple linear regression. The fossils' specimen numbers and species are provided. Because of their varying states of preservation, each individual's recorded landmarks are listed. The corrections are performed for chimp (C), gorilla (G) and human (H) reference samples, and the mean of their residuals are reported. The table is continued in table 5 on the facing page. The weighted mean residuals of the crania presented here and in table 5 are given under "Total Means" on that table. Note that regardless of the genus, the regression-based method using the (large) human reference sample outperforms all other methods.

Specimen	Species	No. of Landmarks	Recorded Landmarks	Mean Subst			TPS			Regression		
				C	G	H	C	G	H	C	G	H
KNM-WT 17400	<i>Paranthropus boisei</i>	7	NA, NSL, IS, FMN left, FMN right, MT left, MT right	0.980	1.140	1.162	1.563	1.911	1.590	0.891	0.965	0.783
Taung	<i>Australopithecus africanus</i>	11	NA, NSL, IS, FMN left, FMN right, ZI left, ZI right, FM right, ZTS right, MT left, MT right	0.450	0.569	0.468	0.453	0.567	0.530	0.455	0.341	0.330
KNM-WT 15000	<i>Homo erectus</i>	10	NA, IS, FMN left, FMN right, ZI left, ZI right, FM left, FM right, ZTS right, MT left	0.990	1.090	0.935	1.614	1.471	0.622	0.936	1.070	0.803
KNM-WT 17000	<i>Paranthropus aethiopicus</i>	10	NA, NSL, IS, FMN left, FMN right, ZI left, ZI right, FM left, MT left, MT right	1.652	1.813	1.930	1.909	1.860	0.829	3.282	1.244	1.086
Total Means				1.083	1.212	1.156	1.357	1.389	1.134	1.684	1.168	0.809
Method Means					1.150			1.294			1.221	

Table 5: Results of landmark estimation on fossil specimens, continued from table 4 on the facing page. "Total Means" gives the weighted mean residuals for the crania here, and those on table 4. "Method means" calculates a weighted mean of residuals grouped by reconstruction method.

Specimen	No. Landmarks	Mean Substitution	Thin Plate Splines
STS 5	—	—	—
KNM-ER 1470	7	1.524	2.325
KNM-ER 1813	8	0.977	1.251
KNM-ER 3733	11	1.221	1.258
KNM-ER 406	12	1.143	1.492
KNM-WT 17400	7	1.105	3.251
<b>Taung</b>	<b>11</b>	<b>0.544</b>	<b>0.499</b>
KNM-WT 15000	10	1.135	1.616
KNM-WT 17000	10	1.410	1.524
Total Means		1.116	1.548

Table 6: Corrections of fossil material using STS 5. “Total Means” is the weighted mean of the residuals obtained for each technique. Note the small residuals obtained for Taung. Taung is the only individual here of the same species as STS 5.

Specimen	No. Landmarks	Mean Substitution	Thin Plate Splines
STS 5	12	0.960	1.103
KNM-ER 1470	7	1.601	2.076
KNM-ER 1813	8	0.806	1.052
KNM-ER 3733	12	1.364	1.408
KNM-ER 406	—	—	—
<b>KNM-WT 17400</b>	<b>7</b>	<b>1.074</b>	<b>1.685</b>
Taung	11	0.798	0.951
KNM-WT 15000	10	1.085	1.300
<b>KNM-WT 17000</b>	<b>10</b>	<b>1.367</b>	<b>1.628</b>
Total Means		1.121	1.358

Table 7: Corrections of fossil material using KNM-ER 406. “Total Means” is the weighted mean of the residuals obtained for each technique. The residuals obtained for the individuals listed in bold are surprisingly high, given that they are of the same genus as KNM-ER 406 (*Paranthropus*).

Specimen	No. Landmarks	Mean Substitution	Thin Plate Splines
STS 5	11	1.125	1.207
<b>KNM-ER 1470</b>	<b>6</b>	<b>1.475</b>	<b>1.372</b>
<b>KNM-ER 1813</b>	<b>7</b>	<b>0.969</b>	<b>1.362</b>
KNM-ER 3733	—	—	—
KNM-ER 406	12	1.5150	1.570
KNM-WT 17400	7	0.962	1.281
Taung	10	0.542	0.578
<b>KNM-WT 15000</b>	<b>9</b>	<b>0.942</b>	<b>0.776</b>
KNM-WT 17000	10	1.770	1.753
Total Means		1.204	1.285

Table 8: Corrections of fossil material using KNM-ER 3733. “Total Means” is the weighted mean of the residuals obtained for each technique. The individuals in bold are of the same genus as KNM-ER 3733 (*Homo*).

# Chapter 7

## Discussion and Recommendation

The previous chapter reports the results of our testing regime. This chapter discusses their relevance, including how they answer the research questions posed in chapter 1. It compares our results to those obtained by others, and ends with recommendations (section 7.5, and summarised in a flow chart on page 127) for researchers employing these techniques.

### 7.1 The Analyses

#### 7.1.1 Analysis *I* through *III*

Analyses *I* through *III* (pages 91–94) each correct the same test samples using one of the three correction methods. In the results one can see a standard and expected pattern: within-species correction outperforms across-species correction. The mean residual for within-species human correction via the regression-based method is only slightly larger than 2 mm.

For all but the largest ( $n = 280$ , human) reference sample, the regression-based method produces — statistically significant — larger residuals when performing across-species corrections (in the tables, the results reported off the diagonals). However, when using the 280 individual reference sample, the regression-based method outperforms the other techniques, and so this across-species pattern of poor performance appears to be due to using too small a reference sample.

Thin plate spline and mean substitution corrections are essentially equivalent for across-species correction, while thin plate splines outperforms mean substitution for within-species corrections. These analyses also suggest that both methods are invariant to increases in reference sample sizes (no significant difference between means as the reference sample sizes increase,  $P > 0.05$ ), which is supported by later analyses.

In relation to within-species correction, the regression-based method outperforms mean substitution even when using a smaller reference sample of only 57 individuals. Within-species, the thin plate spline technique appears to be no different in performance to the regression-based method for all but the largest reference sample.

This would indicate that the thin plate spline method is the best performer, obtaining the lowest total mean, while multiple linear regression has obtained the largest. However, the regression method outperformed thin plate splines in all cases using the 280 individual human reference sample, suggesting that the method's poor performance was due solely to reference samples being too small.

### 7.1.2 Analysis IV

Analysis *IV* determines at which point, relative to reference sample size, the regression-based method outperforms the other techniques. The analysis clearly demonstrates the method's need for large reference samples: the larger the samples, the smaller the residuals. It demonstrates the existence of an asymptote: there is a point after which increasing the reference sample size gives negligible returns in residual reduction.

As in the previous three analyses, we can see a relative invariance in both the mean substitution and thin plate spline methods to changes in the reference sample sizes.

The analysis shows that with a large enough reference sample, multiple linear regression outperforms both thin plate splines and mean substitution. This is likely due to the regression coefficients making use of the landmarks' variation / covariation pattern, something that neither of the other techniques do.

Table 16 shows that, for within-species correction, multiple linear regression outperforms mean substitution from about 50 reference individuals onwards; at between 50 and 100 reference individuals it also begins to outperform thin plate splines (table 17). The differences between the means at these points, and onwards, are significant at  $P < 0.05$ .

Concerning across-species correction, the regression equations only come into their own when using reference samples of some few hundred individuals. From between 250 and 300 individual onwards, multiple linear regression outperforms mean substitution on both the chimp and gorilla test samples. Fewer individuals are required to outperform thin plate splines; this occurs from between 200 and 250 individuals onwards (all means significantly different with  $P < 0.05$ ).

The performance of the mean substitution and thin plate spline methods is also compared using small reference samples of between one and ten individuals (figure 19). The results are similar to those obtained when using larger reference samples, with neither noticeably outperforming the

other. Also, because this small reference sample analysis makes use of reference samples drawn from various species, we can see that thin plate splines outperform mean substitution for within-species correction (as previously seen), and they almost always outperform mean substitution when gorilla is used as the reference sample. Perhaps this is due to gorilla showing great within-species variability, due to their large patterns of sexual dimorphism. This could be something that thin plate splines are less effected by due to their interpolation of the consensus form produced from such variable data.

One important comparison is that between across-species correction of a test sample and within-species correction of the same sample. This is because across-species correction allows us to make use of large reference samples (possibly hundreds of individuals) of extant species, as compared with small reference samples of extinct species (sometimes only a single individual). As we are already reconstructing test samples using large human reference samples, the analysis was extended to correct these same samples using mean substitution and thin plate splines driven by small reference samples of chimp and gorilla.

We then compare the means obtained by correcting the chimp and gorilla test samples with regression equations (driven by the large, 600 individual, human reference sample), against the means obtained correcting these test samples with mean substitution and thin plate splines (driven by within-species reference samples). Table 24 shows *t*-test comparisons between the means obtained by these methods. As before, thin plate splines outperforms mean substitution for within-species correction. The across-species regression method is unable to outperform within-species mean substitution and thin plate splines. In this case, though, the regression method shows no significant difference to mean substitution when correcting the chimp test sample, even though mean substitution has the advantage of using a within-species reference sample. However, when all three techniques use an across-species reference sample, the regression-based method produces the smaller residuals in all but one case: against mean substitution correcting gorilla. Still, the obtained means are not significantly different. This suggests two things: *if even a small, within-species reference sample is available, thin plate splines should be the method of choice. However, if no such sample is available, or if the species of the damaged individual is uncertain, a regression-based technique using a large, across-species reference sample should be employed.*

A *t*-test comparison between the means obtained using mean substitution and thin plate splines is also carried out. The results show a similar pattern to those obtained in previous analyses (compare tables 23 and 13).

### 7.1.3 Analysis V

It is important to consider the amount of damage requiring correction; the more damage, the less undamaged morphology is available for use in the correction. Analysis V (section 6.1.5) examines how the various methods react as the number of missing landmarks requiring estimation increases.

In figure 20 we can see that both mean substitution and multiple linear regression show a fairly low rate of residual increase relative to thin plate splines. Significance tests between the means obtained by each method show that the thin plate spline technique produces the largest residuals as the number of missing landmarks increases. Indeed, from four missing landmarks upwards mean substitution produces lower residuals — or means of no significant difference with  $P > 0.05$  — than thin plate splines. The regression method is essentially identical to thin plate splines when estimating one missing landmark; from two landmarks upwards, multiple linear regression produces the lower mean. Compared to mean substitution, the regression-based method produces a lower mean from the first missing landmark onwards.

Mean substitution appears fairly invariant to the number of missing landmarks, producing a significantly different mean only after a large number require estimation (nine or ten — roughly 70% – 75% of the landmarks). Both the thin plate spline and regression-based methods do not react as well to the number of missing landmarks, probably due to both techniques requiring undamaged morphology to guide the correction — either as predictor variables, or in defining the warp used to estimate the missing landmarks. Both the regression-based method and thin plate spline warps can produce significantly larger residuals from 2 missing landmarks onwards.

The residuals associated with the thin plate spline method grow far more rapidly than those of the other techniques. This may be due to a property of thin plate splines themselves: once defined, the deformation applied by a such a spline does not gradually approach zero the further one moves from the subset of the spline's domain<sup>1</sup> used in its construction; rather the spline (when visualised as a warped plate) becomes increasingly “flat” as one moves away from defining subset, although not constant, and not necessarily zero (Bookstein, 1989). This has the effect that the further one moves from any of the non-missing landmarks used in its construction, the less the spline reflects the required warp in the damaged area. As a greater percentage of landmarks require estimation, the distance between the missing and non-missing landmarks increases, thereby exacerbating this effect. Clumping of correct landmarks (such as with Stw 505, for which only one half of the cranium exists), or similarly, a large area with few landmarks, may indicate that a thin plate spline method is not ideal for reconstructing that particular cranium’s landmarks.

The thin plate spline’s pattern of a rapid residual increase remains true even when small reference samples are used (as in figure 21). Notice that, as with previous analyses, thin plate splines

---

<sup>1</sup>Its input domain: a thin plate spline is merely a function.

produce smaller residuals when using the gorilla reference sample, but this appears to be the only case in which thin plate splines would be chosen over, say, mean substitution, when four or more (roughly 30%) of the landmarks are missing, and this only when correcting a non-gorilla individual.

#### 7.1.4 Analysis VI

Because of the regression-based method's requirement for larger reference samples as more landmarks are involved in a reconstruction, it is interesting to consider which landmarks contribute the most error in a reconstruction. This information comes in useful when the number of reference individuals available to the researcher is small enough to limit the number of landmarks that can be involved in the reconstruction. When attempting to reduce the size of the required reference sample, the landmarks typically associated with the largest errors could be the first considered for removal from the analyses. This will make the solution tractable while removing those landmarks for which less information can be obtained.

Analysis VI, and figures 22(b), 23(b) and 24(b), show how the residuals are distributed over the landmarks used in this thesis. It is clear that for all three techniques the non-facial landmarks contribute the most error, with facial landmarks giving residuals smaller on average by between 1 mm and 2 mm, or 20% for mean substitution, 37% for thin plate splines, and 26% for multiple linear regression.

While the reason for this is not particularly clear, it does suggest that the various techniques perform better reconstructing closely spaced landmarks, such as the facial landmarks in the data set used here. We have already seen a suggestion of this for the thin plate spline method in the previous section. It also suggests that if there is a need to reduce the number of landmarks involved in a reconstruction, then the more sparsely spaced landmarks should be removed first. However, this relation between sparse landmarks and increased residual size is unclear, and further work is required to demonstrate such a relationship. One should also note that the distance between landmarks discussed here is a distance between those landmarks *requiring correction*; the distance between landmarks discussed in the previous section is a distance between those landmarks requiring correction *to those which do not*.

#### 7.1.5 Analysis VII

All the previous analyses are carried out on living species. In contrast, this analysis reconstructs fossil specimens, as reported in tables 4 to 8 on pages 110–112.

Tables 4 and 5 report on across-species reconstruction of fossil crania using living species; while there is no significant differences between the means obtained by the techniques as a whole, multiple linear regression driven by the large human reference sample significantly outperforms the other methods ( $P < 0.05$ ), producing a residual that is between 2 mm and 8 mm smaller. This result is in line with those presented in analysis IV, concerning across-species reconstruction. Table 3 presents mean residuals for correcting these fossils calculated for correcting the specimens grouped into australopith and *Homo* groups. Chimp and gorilla reference samples can be expected to produce smaller residuals than the human reference sample for the australopith group of specimens, due to chimps and gorillas being morphologically closer to australopiths than humans are. However, the results show the using the regression-based method with the large human reference sample produces smaller residuals than the other reconstruction methods and reference samples. This implies that reference sample size, and presumably the associated robust model of variance / covariance structure, is more important than morphological similarity. This result is unexpected.

Tables 6, 7 and 8 correct the fossil specimens using a case typically seen in the literature: a single reference individual of an extinct species. While the Taung child generally reacts well to reconstruction using these samples, across-species correction via multiple linear regression using a large reference sample outperforms these techniques ( $P < 0.05$ ), producing a mean smaller by between 2 mm and 7 mm. Regression equations driven by a human reference sample is the only technique to produce mean residuals below 1 cm. Using an across-species, regression-based method with a reference sample drawn from a living species appears significantly (and greatly) better than using a single, arbitrary reference individual from an extinct species (or even a sample) when the species of the reference individual does not match the species of the damaged individual. Again, this implies that sample size is more important than morphological closeness.

Assuming one is using the regression-based method, larger reference samples of extant species appear more useful than smaller reference samples of a more morphologically similar fossil species. However, it does appear that, if given two large reference samples, the sample of closer morphological affinity to the damaged individual should be chosen. This is suggested in table 3, where the regression equations — using the large human reference sample — corrects *Homo* far better than it does the morphologically less similar australopith specimens.

## 7.2 Observations

Some general observations can be made concerning these results. We begin by considering the research questions put forward in section 1.3.2.

### 7.2.1 The Research Questions

#### 1. Amount of error.

From the analyses, we can see that the error associated with a given technique depends on a number of things. Some landmarks appear to, in general, have larger errors associated with their reconstruction, while within-species correction significantly reduces reconstruction error. Further, the more landmarks there are requiring correction, the greater the error. This is most true when using thin plate splines.

However, some general statements can be made concerning the original question. When performing within-species correction of a single landmark, using reference samples of large enough size, mean substitution does produce residuals larger than thin plate splines, which produce residuals larger than those of multiple linear regression. However, as the number of landmarks increase, thin plate splines quickly produce errors larger than mean substitution (in the results presented here, from 30% and upwards of missing landmarks).

When performing across-species correction, the thin plate spline and mean substitution techniques appear to produce similar sized residuals. Multiple linear regression outperforms both.

#### 2. Growth in error.

As expected, mean substitution shows a greater ability to simultaneously correct multiple landmarks with little increase in error. Both thin plate splines and multiple linear regression display an expected increase in residual sizes. Thin plate splines, however, are surprising in this respect: it appears that these splines may not be appropriate when reconstructing large areas of damaged morphology.

#### 3. Small reference samples.

For within-species correction of a single landmark, mean substitution produces larger residuals than thin plate splines, as expected. As with larger reference samples, the accuracy of thin plate splines quickly deteriorates with an increase in the number of missing landmarks. However, the splines continue to perform well when using the gorilla reference sample, perhaps due to a better ability to handle the large variability within gorillas.

#### 4. Across-species correction.

The thin plate spline and mean substitution methods appear to produce residuals of similar size when used in across-species correction of single landmarks. Across-species correction using these techniques are unable to produce residuals of equal or smaller size than produced by within-species correction with the same techniques. This latter property is true for the regression method as well.

However, regression equations greatly outperform the other techniques when all are used in an across-species fashion; indeed, to such an extant that if the species of a damaged individual is unknown, or there is no within-species sample to use for correction (i.e., it is the sole representative of its species), the regression method, using a large, across-species reference sample, should be preferred.

### 7.2.2 Reference Sample Sizes

Both the mean substitution and thin plate spline methods show no noticeable change in their residuals when using small or large reference samples, except in one case: mean substitution supplies smaller residuals for the within species correction of humans when using more than one reference individual (table 33). However, this could just be an artifact of the small reference sample, and the analysis should be repeated independently to confirm this. If this is the case, it only stresses that while the mean substitution and thin plate spline methods show little reliance on reference sample size, a badly chosen reference individual or small sample can unduly skew the reconstruction. Single reference individuals and small samples should be avoided wherever possible.

Accuracy of the regression method is clearly dependant on the size of the reference sample. There are circumstances in which the regression method should be employed (such as when the species of the damaged individual is unknown), and then a large reference sample must be used.

### 7.2.3 Selecting Landmarks

Much has already been said concerning the selection of landmarks in section 7.1.4. To repeat, if the researcher has a choice of landmarks to exclude from a reconstruction, the results obtained in analysis VI (section 6.1.6) suggest that the landmarks of sparsely spaced regions, such as the neurocranium in this example, should be the first considered for removal.

Due to the nature of the regression method, not only landmarks may be removed from a reconstruction, but also distances between landmarks (see section 5.2.3). The results of previous analyses, when combined with the representation of landmarks as distances, suggest again that those distances associated with sparsely landmarked regions should be considered first for removal.

### 7.2.4 The Importance of Morphological Distance

Zollikofer and Ponce de León (2005) note that using a sample of individuals to infer missing anatomy is common. Indeed, their work is a prime example of the technique, as is the reference based correction method of Gunz (2005).

Zollikofer and Ponce de León (2005) state that when using “extrinsic” information (i.e., information determined from outside of the specimen under consideration), such as a reference sample, we should presuppose that:

- The fossil belongs to the reference sample we are using for correction.
- The information used to perform the correction is based on homologous anatomy.

The authors continue to supply good advice: “... the comparative sample [our reference sample] must represent the shared ancestral pattern of variation rather than patterns of variation characteristic of the derived taxa.” (Zollikofer and Ponce de León, 2005, p. 179). This is in order to avoid biasing the reconstructions towards “preconceived morphologies”, although one may ask if this advice would not bias a derived form towards the more ancestral form.

Their advice may represent the ideal case, but there is a difficulty in applying it to fossil specimens. We only have extremely small samples of fossil material from various hominin species. These samples are small enough that robust statistical inference (and, indeed, robust inferences in general) and reconstruction become difficult. This lack of samples applies not only to members of the same species (if more than one individual is even known), but to the supposed ancestral species as well. This means that if we want to use extrinsic, comparative samples to perform anatomical reconstructions we may not be able to:

1. draw the reference sample from the same population to which the damaged individual belongs,
2. or draw the reference sample from an ancestral population.
3. Further, we may not even be able to create a large enough sample of related individuals, whether or not they share the derived form, the ancestral form, or have autapomorphies (a unique character derived in that species) of their own.

The solution that we supply, as suggested by various studies of variation / covariation in taxa within the primate order (as with Ackermann, 2002, 2005; Ackermann and Cheverud, 2000, 2002) is to use extant species as the comparative / reference sample. Of course this is not ideal. There are differences in variance and covariance patterns among species, and the consequences of such differences are demonstrated by the results presented and discussed in this and the previous chapters. For instance, across-species correction of an individual performs worse than a within-species correction of the same individual. However, the results show that, reservations of biasing the reconstruction aside, using large samples drawn from living species in an across-species reconstruction via regression equations is a much better choice than an across-species correction using small samples of

extinct species (via thin plate splines). In other words, morphological similarity appears to be of less importance than large reference samples.

Ultimately, the reconstruction process is a biasing processes. Those using reconstructions should bear this in mind, and not rely on any one single reconstruction.

### 7.3 Comparison to other results

Gunz (2005) supplies some comparison between techniques in chapter five of his thesis. As in this work, Gunz compares three techniques: mean substitution, thin plate splines, and a regression based method.

All these methods are tested on a sample of 52 *H. sapiens* crania obtained using a Microscribe G2X contact digitiser. Each individual is represented by 388 landmarks and semilandmarks. Portions of each cranium are removed and then estimated using each of the techniques. The author does not describe the reference sample used to perform the reconstruction, and the results are presented as a bar graph of squared residuals with no numerical information. These results suggest<sup>2</sup> that the regression-based method is the best performer, followed by thin plate splines and, finally, mean substitution. However, mean substitution is shown to never outperform the other methods. This is at odds with the results presented here. It can be assumed, however, that Gunz made use of a reference sample large enough to generally guarantee the regression method's superiority given multiple landmarks requiring simultaneous estimation. Table 16 shows that it is clearly possible for the regression method to outperform mean substitution for within species correction using a reference sample of only 50 individuals (if his regression method performs as the one presented here, Gunz could only obtain a reference sample this large if he used his testing sample as his reference sample, given the number of individuals listed in his work), partly explaining this result. However, Gunz's reported performance difference between the thin plate spline and mean substitution techniques is not always borne out by the results presented here, especially when considering the splines' rapid increase in residual size.

Gunz reports that his thin plate spline method outperforms his regression-based method in only one case, that involving a small missing portion of the cranium. Gunz (2005) ascribes this to thin plate splines being ideal for correcting areas on the smooth neurocranium. This result is in line with those presented here in that thin plate splines appear ideal for correcting only a small number of landmarks, especially when using a within-species reference sample as in Gunz's work.

<sup>2</sup>With no numerical information, and no indication of the make up of the reference sample, these results can only be suggestive.

Our results	Gunz's (2005) results
<i>Regression</i> method is the best performer.	<i>Regression</i> method is the best performer.
<i>TPS</i> outperforms regression when estimating small amounts of landmarks using a small reference sample.	<i>TPS</i> outperforms regression when estimating small amounts of landmarks. There appears to have been no testing for the effects of reference sample size on the regression method, and the reference sample size is not stated.
<i>Mean substitution</i> can, given the correct circumstances, outperform the other techniques	<i>Mean substitution</i> is always the worst performing technique

Table 9: A comparison between the results of (Gunz, 2005) and those presented here.

To summarise the difference in results (table 9): ours agree with Gunz's (2005) in that the regression method outperforms the other techniques (although, due to a lack of adequate fossil reference individuals, Gunz does not use the regression-based method for the reconstructions performed in his thesis — he does not use reference samples drawn from extant species as done in our work). We also have thin plate splines outperforming the regression technique in a special case, which Gunz (2005) associates with the correction of the smooth neurocranium, but which our results indicate is more likely due to the area under correction being small (hence reducing the size of the residuals obtained using thin plate splines). Of course, this must also be combined with a small enough reference sample. The largest difference is with Gunz's results and opinions concerning mean substitution: while we show cases in which mean substitution should be used as the corrective method of choice (estimation of many landmarks when only a small reference sample is available), Gunz's results suggest that mean substitution should always perform more poorly than the other techniques, and that it should generally be avoided. The results given here suggest that mean substitution has a place in virtual reconstruction, in certain situations.

## 7.4 Problems

Analysis V assumes that the probability of a specimen missing a landmark is independent of it missing any other (i.e., missing landmarks are chosen randomly). This is not a valid assumption; taphonomic distortion effects whole areas of a specimen, leading to a pattern of damage in which landmarks lying close together are more likely to be affected than those randomly scattered over the specimen's surface. A random scatter should be easier to correct than whole areas, since much of the surrounding context remains intact for use in regression or warping based landmark corrections.

However, as the number of damaged landmarks increases, randomised missing landmarks begin to approximate area (i.e., dependent) damage. This is due to there being a finite number of landmarks on a given specimen, so many of the missing landmarks will lie beside one another. In total, this means the residuals for analysis V generally underestimate those seen in any damaged specimens we may wish to study: the reader should expect to obtain *at least* the error reported here.

As a special case, mean substitution is invariant to the probability of the damage being independent or not, since it makes no use of morphological context. While it does perform an initial alignment between the consensus and damaged forms, as long as the only undamaged landmarks are not clustered in a single, small portion of the specimen<sup>3</sup>, the dependence / independence of landmark damage will not effect the alignment.

Studying a technique's cumulative errors in relation to the number of missing landmarks is perhaps deceptive. Most of the techniques (all except mean substitution) have a reliance on the distance to the closest non-missing landmark; this is arguably a stronger determinant of a correction's accuracy. Even with thin plate splines — which shows high rates of error accumulation — the correction of a missing landmark close to a non-missing landmark (i.e., a landmark used to define the splines) in the consensus form should show less error than those landmarks farther away, even if few landmarks are used to define the spline warp. However, while one would like to hypothesise that these distance-effects result in the greater error associated with the neurocranium's landmarks, mean substitution also displays increased errors with reconstructed neurocranium landmarks, even though mean substitution should not display these 'distance-to-the-closest-non-missing-landmark effects'.

Most of these analyses should be repeated with larger reference samples from other species. Analyses IV and V would benefit the most from this.

A subtle improvement concerns the properties of the correction techniques themselves. Mean substitution and thin plate spline techniques, as implemented here as coordinate-based approaches, classify a landmark as being either present (i.e., "non-damaged") or missing (i.e., "damaged"); missing landmarks are then estimated from some known data / reference sample. However, methods based on distances allow for the *relationship* between landmarks — rather than the landmarks themselves — to be classified as present or missing. Landmarks in their correct position relative to some — but not all — of the other landmarks may then partially remain in the correction, represented by their related distances. In this case, only some — and not all — of the relationships between such a landmark and all the others need be estimated, thereby strengthening the obtained reconstruction. This state of a landmark being only "partially damaged" should clearly be modelled in future analyses so as to test the utility of such approaches.

---

<sup>3</sup>And, technically as long as they are not clustered in a straight line, which would introduce multiple alignment solutions.

## 7.5 Recommendations

The results highlight points to be kept in mind when performing or analysing a reconstruction. These are summed up in this section as “recommendations” to the reader.

### 7.5.1 On choosing a reference sample

Reference samples should generally be drawn from the same species as that of the individual being reconstructed. This clearly results in a reconstruction with a lower mean residual. Small, within-species reference samples often prove adequate to drive mean substitution and warping methods. This is especially true if the amount of damage to be corrected is small, reducing the residuals associated with warping via thin plate splines. However, small samples (especially “samples” of one individual) are highly effected by the individuals making up the sample; it is easy — perhaps too easy — to unduly effect the reconstruction with a bad sample. So while small samples may prove adequate, they should generally be avoided. This is even more true when only a single reference individual is used; this should be avoided wherever possible. If the reference individual is of well-sampled hypodigms, such as *Homo sapiens*, *Homo neandertalensis*, or *Homo erectus*, there is no reason why more than one individual should not be used.

If the species of the damaged individual is unknown or uncertain, a large reference sample is required. This should be used to drive a regression-based reconstruction. In this case, reference samples of a few hundred individuals drawn from an extant species prove satisfactory.

A reference sample drawn from a species showing great intraspecific variability (such as gorilla) should be avoided. Residuals obtained using such a reference sample appear to be larger than those obtained using other species. If such a species must be used, thin plate splines appear to be a good correction technique to employ.

### 7.5.2 On choosing the technique

The most clear cut recommendation involves correcting individuals of a species for which a large reference sample can be drawn. Regression methods are, in this case, clearly superior to other methods. However, we often cannot find such a large sample, complicating our choice of technique. The following guidelines are suggested by the results of our testing and assume that the researcher does not exclude mean substitution because of its nature (see section 5.2.1).

Researchers should first ask themselves this question concerning the reconstruction: *Is the species of the damaged individual known or unknown?*

- **A known species.**

*Can we create a within-species reference sample?* If no, proceed as if the species of the damaged individual were unknown.

*How large is the within-species reference sample?*

- **The reference sample has 150 individuals or more.**

A regression-based method should be employed.

- **The reference sample is small.**

*Is there much damage, represented by a large portion of the landmarks requiring estimation?*

- \* **There is significant damage.**

Mean substitution should be used unless the reference sample is from a species with large intraspecific variability. Thin plate splines should then be used.

- \* **There is little damage.**

Thin plate splines should be used.

- **An unknown species.**

In this case, across-species correction is indicated. **Draw a reference sample of 300 or more individuals from some extant species that does not show great intraspecific variability (e.g., not gorillas, for example).** A regression-based correction method should be used. Due to the ease with which landmark data for extant species may be obtained, there is little reason not to employ such a correction regime. However, in case this does prove difficult:

*How large a reference sample is available?*

- **The reference sample is larger than 300 individuals.**

This is the case just described: a regression-based approach should be used.

- **The reference sample is smaller than 250 – 300 individuals.**

*Is there much damage, represented by a large portion of the landmarks requiring estimation?*

- \* **There is significant damage.**

Mean substitution should be used unless the reference sample is from a species with large intraspecific variability. Thin plate splines should then be used.

- \* **There is little damage.**

Thin plate splines should be used.

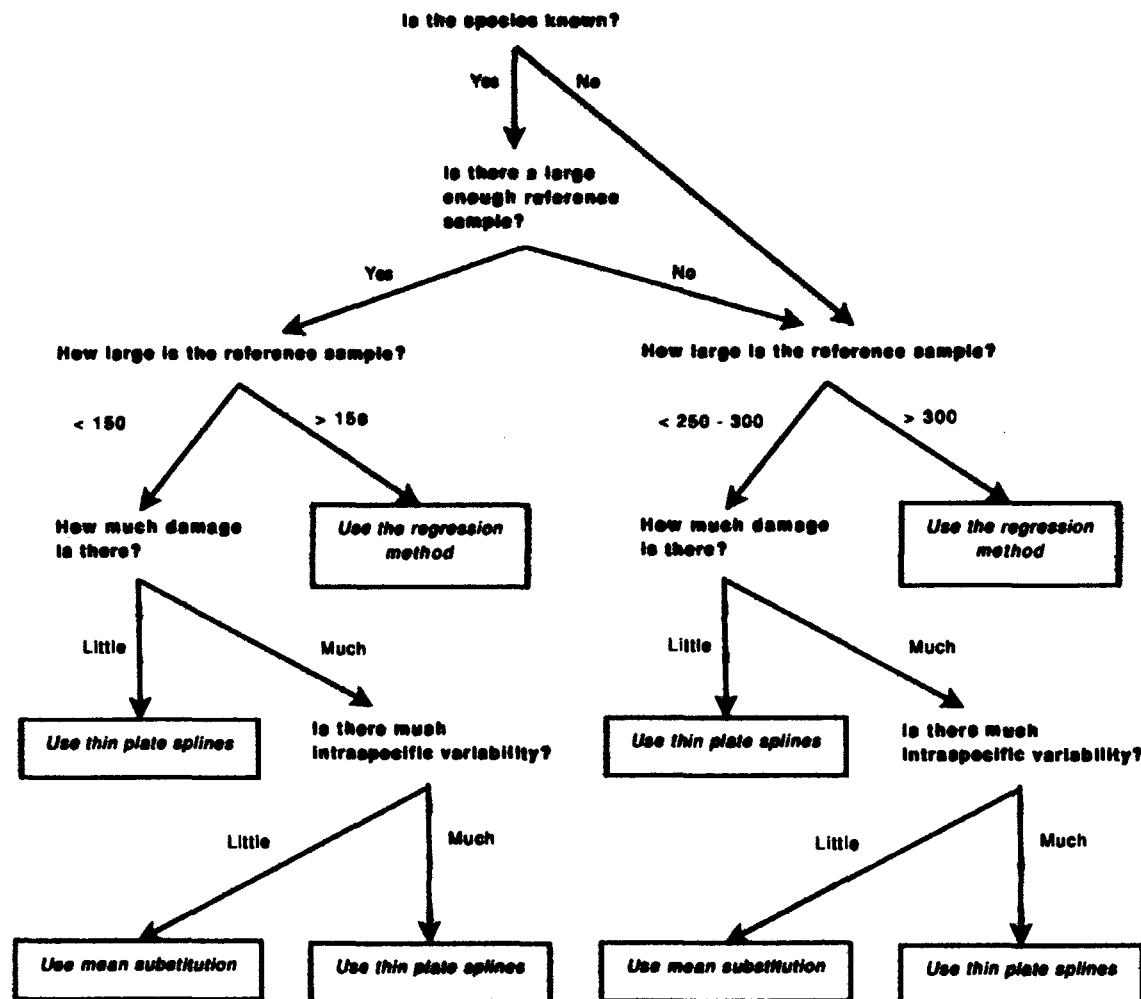


Figure 25: A flow chart outlining the recommendations of section 7.5. These recommendations detail how to best choose a correction technique, given the available reference sample sizes and the amount of damage to be corrected.



# Chapter 8

## Conclusion

In the previous chapters we presented descriptions of the techniques themselves (chapter 5), the testing regime (chapter 6), and a discussion of results (chapter 7). Also covered were questions of data capture (chapter 3) and the display of this data (chapter 4). In this, our final chapter, we summarise the main body of the work. The chapter outlines our aims, discussing how well these have been met. It supplies a brief overview of the testing and the obtained results, followed by the contributions of the thesis. Since fossil reconstruction is a rich and detailed field, drawing on many areas and likely to grow considerably over the next few years, the chapter concludes with an overview of future work.

### 8.1 Summary of the Aims

As given in chapter 1, the main aims of this thesis have been:

- *to empirically compare various statistical (mean substitution; multiple linear regression) and geometrical (thin plate spline) reconstruction methods for fossil material.* This was done by collecting landmark data sets of undamaged individuals. Each landmark was, in turn, removed and estimated using a given correction technique. The landmark estimate could then be compared with the known landmark, and a residual calculated. The resulting residuals were then compared across the different techniques.
- *to examine the effect that reference sample size has on the techniques.* This was done by repeating the above testing procedure while varying the reference sample sizes. Reference samples from only one individual — referring back to the examples of previous work in section 2.7.2, reveals that this is an unfortunately typical case — to up to 600 individuals were employed.

- *to determine if reference samples used to correct fossil specimens can be drawn from extant species.* This was done by varying the species of the reference sample used in the corrections. Residuals obtained for using within-species reference samples were compared with those from across-species samples.

## 8.2 Overview of the Testing and Results

### 8.2.1 Exploratory Testing of the Reconstruction Methods: Analysis I–III

The initial tests concerned correcting samples of chimp, gorilla and human individuals using each of the three reconstruction methods. This was done by sequentially removing and estimating landmarks. The regression-based approach does not appear to fare as well as expected (previous work leads us to believe that the regression method is far superior to the other methods employed here) for all but the largest ( $n = 280$ ) reference sample.

### 8.2.2 When Does Regression Become Viable: Analysis IV

Clearly, the regression-based method requires large reference samples. The next test was designed to examine at what point the reference samples are large enough for the regression method to outperform the other techniques. This was done by repeatedly correcting the same test samples with each of the techniques, while increasing the reference sample size from ten to 600 individuals.

Multiple linear regression always outperforms mean substitution for within-species corrections, from a sample sizes of roughly 50 individuals and above. For small sample sizes, multiple linear regression produces residuals with a mean of no significant difference to that produced by thin plate splines, until sample sizes of roughly 150 individuals are used. From this point onwards, regression equations produce the significantly smaller mean residual.

For across-species corrections, regression equations do not perform with any significant difference to mean substitution when using sample sizes of 200 individuals. From 300 individuals and above it outperforms mean substitution. Compared with thin plate splines, regression performs similarly at across-species reconstruction from 150 individuals, and outperforms it at 250 individuals onwards.

We notice that when mean substitution and thin plate splines use a within-species reference sample, they outperform the regression-based method at correcting a test sample if the regression-based method uses an across-species sample. However, if all the methods make use of an across species sample, regression equations produces the smaller residuals.

### 8.2.3 How Does the Amount of Damage Effect Residuals: Analysis V

All of the testing so far has determined residuals based on correcting only a single missing landmark. This is an unrealistic case, and this analysis involves reconstructing multiple missing landmarks simultaneously. Landmarks are removed in groups from one up to ten landmarks. All the missing landmarks are corrected using the remaining landmarks as input data.

Mean substitution, as expected, varies little in its obtained residuals as the extant of damage increases. Thin plate splines shows a marked increase in residual size, far greater than that of any other technique. The method rapidly becomes a poor choice for reconstruction work, with the other techniques producing smaller residuals from, in general, four missing landmarks and upwards.

### 8.2.4 Do Landmarks Benefit Equally From Reconstruction: Analysis VI

This analysis reconstructed landmarks over different parts of the cranium. Landmarks were partitioned into facial and non-facial landmarks, with mean residuals calculated for both groups. Irrespective of method, the mean residuals obtained for the non-facial landmarks were always significantly larger than for the facial landmarks. It appears that all landmarks are not reconstructed equally; this may be an artifact of landmark density.

### 8.2.5 How Well Do Fossil Specimens Fare Under Reconstruction: Analysis VII

While the previous analyses were carried out on living species (to allow for large enough numbers for both the test and reference samples), this analysis performed reconstruction of known fossil specimens. The reconstructions were performed using all three reconstruction techniques, and reference samples drawn from living species. The reconstructions were repeated using a single fossil specimen as a reference individual, and mean substitution and thin plate splines as the correction method. The best overall correction method was multiple linear regression using a large ( $n = 628$ ) human reference sample. Since the species of the reference and test individuals did not necessarily match, this outcome was partly expected given previous analyses; however, the use of a large human reference sample outperforming corrections using a fossil reference of the same genus as the damaged individual was surprising. Out of all of the method / reference sample combinations, the regression method / human sample pair was the only combination to produce a mean residual below 1 cm: at least 2 mm, or 20%, better than any other method, even when fossil reference individuals were employed.

### 8.3 Contributions

We have performed the following in the completion of this work:

- While data was obtained from various sources (contact digitisers, computed tomography), we directly applied a photogrammetric pipeline to obtain surface data from fossil hominin specimens, as described in chapter 3. This included the determination of target positions on a reference frame; the photography of fossil crania; camera calibration, followed by the identification and triangulation of interest points across multiple photographs of each individual cranium.
- Various data processing and rendering algorithms were explored, as outlined in chapter 4. This included surface construction via the powercrust algorithm (a Delaunay based algorithm), and marching cubes. Point set rendering methods were also investigated.
- The novel contribution of this thesis was the testing regime of chapters 6 and 7. The various tests undertaken have already been outlined in this chapter (section 8.2). While other authors have reported some results concerning the capabilities of their reconstruction methods, and even compared various techniques, this thesis contributes a more rigorous examination of these methods.
- An important outcome of this thesis is a set of recommendations for the use of the mean substitution, thin plate spline and regression-based methods. These recommendations follow from the results of the testing, and are given in section 7.5 on page 125.

The important findings of this thesis can be summarised as follows:

- The smallest mean reconstruction error was obtained by within-species correction of *Homo sapiens* via the multiple linear regression method. This achieved a mean residual of 0.19 cm.
- The smallest mean error for the correction of fossil individuals was obtained on crania from various species of *Homo*, reconstructed using multiple linear regression and a large ( $n = 628$ ) modern human reference sample. The mean residual was 0.666 cm.
- When using reference samples too small for regression to be viable, thin plate splines produce the smallest mean residuals for within-species reconstruction of small areas (typically less than 30%).
- In support of this, when using reference samples too small for regression to be viable, mean substitution produces the smallest mean residuals for within-species reconstruction of large areas (typically more than 30%).

- When only across-species reconstruction (of small areas) is possible, the regression-based method produces the smallest residuals (assuming large enough reference samples), while mean substitution and thin plate splines appear to be of equal ability. Thin plate splines again performs poorly for the reconstruction of large areas.
- The amount of damage to be corrected has little effect on the resulting mean residual obtained using mean substitution. The thin plate spline method shows rapid degradation in its performance, to the point that mean substitution can outperform thin plate splines in cases involving the reconstruction of large areas. With large enough reference samples, the regression-based method outperforms the other methods at all levels of damage.
- Reference sample size has little effect on mean substitution and thin plate spline warping methods, except in that larger reference samples provides greater protection against outliers. For this reason, multiple individuals (even if only a few) should be used even for mean substitution or thin plate spline methods.
- The regression-based method can require up to a few hundred individuals (300 or more) for it to outperform the other methods.
- As expected, within-species correction generally outperforms across-species correction.
- Reconstruction of fossil specimens using reference samples drawn from living species is an appropriate choice if the species of the damaged specimen is unknown.
- Large reference samples and the regression method (and presumably the associated robust model of variance / covariance structure) should generally be chosen over small reference samples of individuals of closer morphological affinity to the damaged individual. Morphological affinity only becomes of interest when the reference samples are of similar size.
- Reconstruction using reference individuals drawn from a species with large intra-specific variability (such as those which show high levels of sexual dimorphism) should be avoided, as this produces poorer reconstruction results.

## 8.4 Future Work

1. The testing could be improved in a number of ways:
  - Various problems with the testing methodology have been listed in section 7.4. These should be addressed.

- For many of our studies we had reference samples of chimp and gorilla that were too small to properly test the regression method. In these instances we had to rely on the human test sample. The tests should be repeated using larger reference samples, and incorporating a greater variety of species.
  - Testing should be repeated using semilandmarks, to note how this may effect the various techniques. The results of Gunz (2005) suggest that they may behave differently.
  - Cumulative error tests (analysis V of chapter 6) should have landmarks removed from the same area rather than randomly, so that the test no longer underestimates the error.
  - Richtsmeier et al. (1992) suggests the use of projection pursuit regression (PPR) — a non-parametric regression technique — for statistical reconstruction, rather than linear techniques. It would be interesting to see how well PPR, and other non-parametric regression techniques in general, behaves when applied to fossil reconstruction. Härdle (1990) supplies a general introduction to non-parametric techniques.
2. The warping of one surface onto another is usually carried out using thin plate splines (such as the warping of a portion of the modern human cranium onto the Neandertal le Moustier 1 cranium by Ponce De León and Zollikofer, 1999). However, our test results show that thin plate splines have a rapid growth in residual size, suggesting that thin plate splines may not be the best warping technique for landmark fitting. One wonders if a warping technique which effects only a local area (as opposed to thin plate spline's global effect) could replace thin plate splines as the method of choice, especially if, as with thin plate splines, it can be represented purely in terms of linear algebra (Gunz (2005) and Gunz et al. (2005) argue for the advantages of such representations). The first technique that comes to mind is the *direct manipulation of free form deformation* (Hsu et al., 1992), although more appropriate techniques exist (see Bechmann, 1994; Milliron et al., 2002, as a review of techniques, and a framework for warping, respectively). The techniques could be tested using metrics that measure the distance between surfaces, such as the Hausdorff metric used by Cignoni et al. (1998).
  3. Much virtual and morphometric reconstruction occurs on CT data. While surface models are usually extracted from such voxel data, it would be useful to extend the various techniques provided here, such as the warping techniques, to operate directly on the voxel data.
  4. Plastic taphonomic distortion due to the weight and shifting of sediments is not completely arbitrary. While it remains difficult to determine the forces that damaged a fossil cranium, it might still be useful to investigate warping techniques that simulate physical processes in

correcting these effects (e.g., mass spring models, finite element models, approximate continuum models, low degree of freedom models, etc.; see Gibson and Mirtich, 1997, for a review of such techniques).

5. Often a single specimen is used as a reference model in thin plate spline based reconstructions, sometimes because only a few known members of the species exist. One wonders if it is possible to bulk up a reference sample with individuals from another species, producing a sample consisting of individuals from different species. For instance, is it possible to create a reference sample based around STS 5, but bulked up using chimp individuals. The contribution of STS 5 to the sample should be weighted to increase its importance in the reconstruction.
6. An interesting extension to this work is to automatically identify damaged areas of the cranium. While this appears trivial for missing regions, it becomes more complicated for regions suffering from plastic distortion. Perhaps shape retrieval work can be modified for this application (e.g., Funkhouser et al., 2003, 2005).
7. The coordinate-free regression method used in this work has the advantage that it does not require whole landmarks to be excluded from analyses if only some of the distance data relating it to other landmarks are missing; it is possible to only exclude these distances. An extension of mean substitution and thin plate spline warping methods to operate in a coordinate-free manner seems possible. It would be enlightening to examine how these methods may (or may not) benefit from this. The reconstruction tests should be repeated to take advantage of this ability in coordinate-free approaches. It would also mean a more realistic simulation of taphonomic damage, as distances (rather than landmarks) will be randomly removed.



## Appendix A

### Various Tables

#### A.1 Tables For Analyses *I* through *III*

Test Species	Reference Species		
	Chimp	Gorilla	Human
Chimp	$P = 1.674 \times 10^{-4}$ $t = 5.297$ $df = 12.446$ regression	$P = 0.885$ $t = 0.148$ $df = 10.796$ —	$P = 0.109$ $t = -1.744$ $df = 11.174$ —
Gorilla	$P = 7.384 \times 10^{-5}$ $t = -6.035$ $df = 11.374$ mean subst	$P = 9.576 \times 10^{-8}$ $t = 9.270$ $df = 15.617$ regression	$P = 0.034$ $t = -2.323$ $df = 15.594$ mean subst
Human	$P = 0.030$ $t = -2.546$ $df = 9.414$ mean subst	$P = 4.578 \times 10^{-5}$ $t = -5.620$ $df = 15.257$ mean subst	$P = 9.961 \times 10^{-6}$ $t = 6.560$ $df = 14.708$ regression

Table 10: Two sided  $t$ -test results between the mean substitution (analysis *I*, section 6.1.1) and multiple linear regression (analysis *III*, section 6.1.3) techniques. The results are for reference samples of 57 individuals. For statistically significant differences between the means ( $P < 0.05$ ), the method with the smaller mean is indicated.

Test Species	Reference Species		
	Chimp	Gorilla	Human
Chimp	$P = 1.674 \times 10^{-4}$ $t = 5.297$ $df = 12.446$ regression	$P = 0.200$ $t = -1.379$ $df = 9.432$ —	$P = 2.151 \times 10^{-6}$ $t = 6.920$ $df = 17.444$ regression
Gorilla	$P = 1.208 \times 10^{-6}$ $t = -6.511$ $df = 23.002$ mean subst	$P = 3.965 \times 10^{-6}$ $t = 5.584$ $df = 31.153$ regression	$P = 2.998 \times 10^{-8}$ $t = 6.95$ $df = 37.7$ regression
Human	$P = 9.228 \times 10^{-11}$ $t = -7.789$ $df = 61.958$ mean subst	$P = < 2.2 \times 10^{-16}$ $t = -12.584$ $df = 62.930$ mean subst	$P = < 2.2 \times 10^{-16}$ $t = 30.873$ $df = 117.818$ regression

Table 11: Two sided  $t$ -test results between the mean substitution (analysis I, section 6.1.1) and multiple linear regression (analysis III, section 6.1.3) techniques. The results are for unequal reference samples sizes. For statistically significant differences between the means ( $P < 0.05$ ), the method with the smaller mean is indicated.

Test Species	Reference Species		
	Chimp	Gorilla	Human
Chimp	$P = 0.265$ $t = 1.151$ $df = 17.77$ —	$P = 0.813$ $t = 0.241$ $df = 15.979$ —	$P = 0.798$ $t = -0.262$ $df = 13.474$ —
Gorilla	$P = 1.251 \times 10^{-4}$ $t = -5.275$ $df = 13.734$ tps	$P = 0.051$ $t = 2.150$ $df = 13.132$ —	$P = 0.050$ $t = -2.100$ $df = 17.609$ —
Human	$P = 0.067$ $t = -2.067$ $df = 9.489$ —	$P = 1.845 \times 10^{-5}$ $t = -7.000$ $df = 11.442$ tps	$P = 0.288$ $t = 1.095$ $df = 17.980$ —

Table 12: Two sided  $t$ -test results between the thin plate spline (analysis II, section 6.1.2) and multiple linear regression (analysis III, section 6.1.3) methods. The results are for reference samples of 57 individuals. If the means are significantly different ( $P < 0.05$ ), the technique with the smaller mean residual is indicated.

Test Species	Reference Species		
	Chimp	Gorilla	Human
Chimp	$P = 0.265$	$P = 0.179$	$P = 1.485 \times 10^{-6}$
	$t = 1.151$	$t = -1.432$	$t = 7.829$
	$df = 17.77$	$df = 11.462$	$df = 14.369$
	—	—	regression
Gorilla	$P = 5.837 \times 10^{-6}$	$P = 0.573$	$P = 9.275 \times 10^{-7}$
	$t = -5.513$	$t = 0.569$	$t = 5.977$
	$df = 29.405$	$df = 32.101$	$df = 33.951$
	tps	—	regression
Human	$P = 1.072 \times 10^{-7}$	$P = < 2.2 \times 10^{-16}$	$P = < 2.2 \times 10^{-16}$
	$t = -5.935$	$t = -13.011$	$t = 12.424$
	$df = 68.722$	$df = 66.183$	$df = 104.111$
	tps	tps	regression

Table 13: Two sided  $t$ -test results between the thin plate spline (analysis *II*, section 6.1.2) and linear regression techniques (analysis *III*, section 6.1.3). The results are for unequal reference sample sizes. If the means are significantly different ( $P < 0.05$ ), the method with the smaller mean is indicated.

Test Species	Reference Species		
	Chimp	Gorilla	Human
Chimp	$P = 8.794 \times 10^{-4}$	$P = 0.860$	$P = 0.024$
	$t = -4.269$	$t = 0.181$	$t = 2.491$
	$df = 13.236$	$df = 12.653$	$df = 15.839$
	tps	—	mean subst
Gorilla	$P = 0.497$	$P = 2.097 \times 10^{-4}$	$P = 0.990$
	$t = 0.694$	$t = -4.718$	$t = 0.012$
	$df = 15.952$	$df = 16.636$	$df = 16.869$
	—	tps	—
Human	$P = 0.046$	$P = 0.290$	$P = 1.523 \times 10^{-4}$
	$t = 2.144$	$t = -1.097$	$t = -5.083$
	$df = 17.878$	$df = 14.509$	$df = 14.420$
	mean subst	—	tps

Table 14: Two sided  $t$ -test results between the thin plate spline (analysis *II*, section 6.1.2) and the mean substitution (analysis *I*, section 6.1.1) methods. The results are for reference samples of 57 individuals. If the means are significantly different ( $P < 0.05$ ), the method with the smaller mean is indicated.

Test Species	Reference Species		
	Chimp	Gorilla	Human
Chimp	$P = 8.794 \times 10^{-4}$	$P = 0.747$	$P = 0.034$
	$t = -4.269$	$t = -0.330$	$t = 2.323$
	$df = 13.236$	$df = 12.013$	$df = 15.974$
	tps	—	mean subst
Gorilla	$P = 0.379$	$P = 6.463 \times 10^{-8}$	$P = 0.935$
	$t = 0.893$	$t = -6.695$	$t = 0.082$
	$df = 31.039$	$df = 37.904$	$df = 35.520$
	—	tps	—
Human	$P = 0.152$	$P = 0.049$	$P = < 2.2 \times 10^{-16}$
	$t = 1.488$	$t = -1.988$	$t = -11.847$
	$df = 20.386$	$df = 108.615$	$df = 101.809$
	—	tps	tps

Table 15: Two sided  $t$ -test results between the thin plate spline (analysis *II*, section 6.1.2) and the mean substitution (analysis *I*, section 6.1.1) techniques. The results are for unequal reference sample sizes. If there is a statistically significant difference in the means ( $P < 0.05$ ), the technique with the smaller mean is indicated.

## A.2 Tables For Analysis IV: Cross Over

Table 16: *t*-test results between the mean substitution and regression-based techniques, while increasing the reference sample sizes. If there is a statistically significant difference between the means ( $P < 0.05$ ), the technique with the smaller mean is indicated.

Reference Sizes	Test Species		
	Human	Chimp	Gorilla
50	$P = 0.008$ $t = 2.822$ $df = 31.831$ regression	$P = 4.701 \times 10^{-15}$ $t = -15.192$ $df = 28.015$ mean subst	$P = 4.649 \times 10^{-16}$ $t = -15.120$ $df = 31.748$ mean subst
	$P = 2.307 \times 10^{-10}$ $t = 8.661$ $df = 36.336$ regression	$P = 5.684 \times 10^{-12}$ $t = -11.047$ $df = 29.336$ mean subst	$P = 3.463 \times 10^{-13}$ $t = -12.255$ $df = 29.898$ mean subst
	$P = < 2.2 \times 10^{-16}$ $t = 15.884$ $df = 45.725$ regression	$P = 0.031$ $t = -2.244$ $df = 34.892$ mean subst	$P = 0.110$ $t = -1.635$ $df = 37.919$ —
	$P = < 2.2 \times 10^{-16}$ $t = 20.218$ $df = 48.686$ regression	$P = 0.076$ $t = 1.819$ $df = 42.509$ —	$P = 0.760$ $t = -0.307$ $df = 52.978$ —
250	$P = < 2.2 \times 10^{-16}$ $t = 21.438$ $df = 50.033$ regression	$P = 7.021 \times 10^{-8}$ $t = 6.423$ $df = 45.496$ regression	$P = 0.005$ $t = 2.897$ $df = 53.689$ —
	$P = < 2.2 \times 10^{-16}$ $t = 21.119$ $df = 49.978$ regression	$P = 1.287 \times 10^{-10}$ $t = 8.055$ $df = 50.428$ regression	$P = 7.733 \times 10^{-06}$ $t = 4.947$ $df = 53.975$ regression

Table 16: Continued from previous page.

	Human	Chimp	Gorilla
350	$P = < 2.2 \times 10^{-16}$ $t = 22.832$ $df = 52.667$ regression	$P = 2.437 \times 10^{-10}$ $t = 7.934$ $df = 48.925$ regression	$P = 5.183 \times 10^{-9}$ $t = 6.968$ $df = 52.786$ regression
400	$P = < 2.2 \times 10^{-16}$ $t = 22.321$ $df = 53.527$ regression	$P = 2.366 \times 10^{-11}$ $t = 8.559$ $df = 49.815$ regression	$P = 1.725 \times 10^{-9}$ $t = 7.276$ $df = 52.392$ regression
450	$P = < 2.2 \times 10^{-16}$ $t = 21.489$ $df = 53.066$ regression	$P = 2.809 \times 10^{-15}$ $t = 11.076$ $df = 51.868$ regression	$P = 3.299 \times 10^{-12}$ $t = 9.032$ $df = 51.624$ regression
500	$P = < 2.2 \times 10^{-16}$ $t = 22.006$ $df = 54.000$ regression	$P = < 2.2 \times 10^{-16}$ $t = 14.764$ $df = 53.995$ regression	$P = 6.337 \times 10^{-12}$ $t = 8.844$ $df = 51.719$ regression
550	$P = < 2.2 \times 10^{-16}$ $t = 21.384$ $df = 53.851$ regression	$P = < 2.2 \times 10^{-16}$ $t = 14.768$ $df = 53.993$ regression	$P = 7.442 \times 10^{-14}$ $t = 10.199$ $df = 50.454$ regression
600	$P = < 2.2 \times 10^{-16}$ $t = 22.163$ $df = 53.840$ regression	$P = < 2.2 \times 10^{-16}$ $t = 16.068$ $df = 53.804$ regression	$P = < 2.2 \times 10^{-16}$ $t = 11.61$ $df = 49.788$ regression

Table 17: *t*-test results between the thin plate spline and regression-based techniques, while increasing the reference sample sizes. If the means are significantly different ( $P < 0.05$ ), the method with the smaller mean is indicated.

Reference Sizes	Test Species		
	Human	Chimp	Gorilla
50	$P = 0.062$ $t = -1.920$ $df = 42.324$ —	$P = 1.669 \times 10^{-14}$ $t = -14.093$ $df = 28.991$ tps	$P = < 2.2 \times 10^{-16}$ $t = -14.157$ $df = 34.332$ tps
	$P = 0.119$ $t = 1.587$ $df = 51.091$ —	$P = 1.155 \times 10^{-10}$ $t = -9.388$ $df = 31.653$ tps	$P = 7.557 \times 10^{-13}$ $t = -11.501$ $df = 31.704$ tps
	$P = 1.227 \times 10^{-6}$ $t = 5.472$ $df = 53.145$ regression	$P = 0.879$ $t = 0.154$ $df = 41.588$ —	$P = 0.528$ $t = -0.636$ $df = 43.463$ —
	$P = 7.581 \times 10^{-10}$ $t = 7.559$ $df = 50.488$ regression	$P = 1.078 \times 10^{-05}$ $t = 4.877$ $df = 51.437$ regression	$P = 0.231$ $t = 1.211$ $df = 53.388$ —
250	$P = 9.783 \times 10^{-11}$ $t = 8.17$ $df = 49.494$ regression	$P = 3.465 \times 10^{-12}$ $t = 8.939$ $df = 53.331$ regression	$P = 1.877 \times 10^{-4}$ $t = 4.04$ $df = 49.109$ regression
	$P = 1.535 \times 10^{-10}$ $t = 8.045$ $df = 49.432$ regression	$P = 1.200 \times 10^{-14}$ $t = 10.513$ $df = 53.743$ regression	$P = 4.473 \times 10^{-7}$ $t = 5.774$ $df = 51.56$ regression
	$P = 2.037 \times 10^{-11}$ $t = 8.756$ $df = 46.737$ regression	$P = 5.663 \times 10^{-14}$ $t = 10.053$ $df = 54.000$ regression	$P = 2.590 \times 10^{-9}$ $t = 7.318$ $df = 47.353$ regression

Table 17: Continued from previous page.

	Human	Chimp	Gorilla
400	$P = 3.549 \times 10^{-11}$ $t = 8.627$ $df = 46.047$ regression	$P = 3.613 \times 10^{-14}$ $t = 10.183$ $df = 53.968$ regression	$P = 2.396 \times 10^{-9}$ $t = 7.353$ $df = 46.967$ regression
450	$P = 3.600 \times 10^{-11}$ $t = 8.551$ $df = 47.446$ regression	$P = < 2.2 \times 10^{-16}$ $t = 12.068$ $df = 53.367$ regression	$P = 3.626 \times 10^{-11}$ $t = 8.623$ $df = 46.009$ regression
500	$P = 2.526 \times 10^{-11}$ $t = 8.786$ $df = 45.046$ regression	$P = < 2.2 \times 10^{-16}$ $t = 14.831$ $df = 49.354$ regression	$P = 7.726 \times 10^{-11}$ $t = 8.385$ $df = 46.249$ regression
550	$P = 2.532 \times 10^{-11}$ $t = 8.706$ $df = 46.462$ regression	$P = < 2.2 \times 10^{-16}$ $t = 14.724$ $df = 49.323$ regression	$P = 2.979 \times 10^{-12}$ $t = 9.469$ $df = 44.716$ regression
600	$P = 2.532 \times 10^{-11}$ $t = 8.706$ $df = 46.462$ regression	$P = < 2.2 \times 10^{-16}$ $t = 14.724$ $df = 49.323$ regression	$P = 2.979 \times 10^{-12}$ $t = 9.469$ $df = 44.716$ regression

Table 18:  $t$ -test results between the thin plate spline and mean substitution techniques, while increasing the reference sample sizes. If there is a statistically significant difference between the means, the method with the smaller mean is indicated.

Reference Sizes	Test Species		
	Human	Chimp	Gorilla
50	$P = 4.187 \times 10^{-10}$ $t = -8.121$ $df = 41.436$ tps	$P = 1.261 \times 10^{-4}$ $t = 4.165$ $df = 48.831$ mean subst	$P = 0.165$ $t = 1.409$ $df = 51.524$ —

Table 18: Continued from previous page.

	Human	Chimp	Gorilla
100	$P = 2.421 \times 10^{-10}$ $t = -8.293$ $df = 41.453$ tps	$P = 1.337 \times 10^{-4}$ $t = 4.149$ $df = 48.58$ mean subst	$P = 0.126$ $t = 1.555$ $df = 51.062$ —
150	$P = 3.884 \times 10^{-10}$ $t = -8.094$ $df = 42.368$ tps	$P = 3.154 \times 10^{-4}$ $t = 3.877$ $df = 48.836$ mean subst	$P = 0.159$ $t = 1.431$ $df = 51.201$ —
200	$P = 2.496 \times 10^{-10}$ $t = -8.276$ $df = 41.581$ tps	$P = 9.500 \times 10^{-5}$ $t = 4.255$ $df = 48.532$ mean subst	$P = 0.123$ $t = 1.571$ $df = 51.004$ —
250	$P = 2.058 \times 10^{-10}$ $t = -8.319$ $df = 41.892$ tps	$P = 2.002 \times 10^{-4}$ $t = 4.023$ $df = 48.502$ mean subst	$P = 0.139$ $t = 1.505$ $df = 50.933$ —
300	$P = 3.486 \times 10^{-10}$ $t = -8.160$ $df = 41.773$ tps	$P = 2.013 \times 10^{-4}$ $t = 4.020$ $df = 48.719$ mean subst	$P = 0.151$ $t = 1.457$ $df = 51.106$ —
350	$P = 3.814 \times 10^{-10}$ $t = -8.094$ $df = 42.483$ tps	$P = 6.789 \times 10^{-4}$ $t = 3.629$ $df = 48.899$ mean subst	$P = 0.211$ $t = 1.266$ $df = 51.141$ —
400	$P = 6.313 \times 10^{-10}$ $t = -7.887$ $df = 43.529$ tps	$P = 0.003$ $t = 3.100$ $df = 49.207$ mean subst	$P = 0.314$ $t = 1.016$ $df = 51.350$ —
450	$P = 9.594 \times 10^{-10}$ $t = -7.745$ $df = 43.868$ tps	$P = 0.008$ $t = 2.750$ $df = 49.406$ mean subst	$P = 0.439$ $t = 0.780$ $df = 51.462$ —

Table 18: Continued from previous page.

	Human	Chimp	Gorilla
500	$P = 1.39 \times 10^{-9}$	$P = 0.017$	$P = 0.496$
	$t = -7.586$	$t = 2.469$	$t = 0.686$
	$df = 45.002$	$df = 49.603$	$df = 51.547$
	tps	mean subst	—
550	$P = 1.763 \times 10^{-9}$	$P = 0.023$	$P = 0.530$
	$t = -7.514$	$t = 2.355$	$t = 0.632$
	$df = 45.042$	$df = 49.616$	$df = 51.577$
	tps	mean subst	—
600	$P = 2.239 \times 10^{-9}$	$P = 0.033$	$P = 0.595$
	$t = -7.424$	$t = 2.191$	$t = 0.535$
	$df = 45.552$	$df = 49.776$	$df = 51.651$
	tps	mean subst	—

Chimp Reference Sample					
Method	Reference sample size				
	1	2	3	4	5
<i>Chimp Test Sample</i>					
P	$6.294 \times 10^{-10}$	$1.630 \times 10^{-9}$	$3.238 \times 10^{-7}$	$1.445 \times 10^{-7}$	$1.388 \times 10^{-9}$
t	-7.904	-7.521	-5.903	-6.272	-7.71
df	43.196	45.457	49.291	43.204	42.185
Result	TPS	TPS	TPS	TPS	TPS
<i>Gorilla Test Sample</i>					
P	0.115	0.929	0.018	0.055	0.090
t	-1.603	-0.09	2.444	1.963	1.729
df	50.87	51.959	49.645	49.461	49.499
Result	—	—	Mean Subst	—	—
<i>Human Test Sample</i>					
P	$5.761 \times 10^{-8}$	$1.357 \times 10^{-5}$	$5.415 \times 10^{-4}$	$1.395 \times 10^{-5}$	$3.833 \times 10^{-5}$
t	6.356	4.859	3.703	4.837	4.523
df	50.772	46.882	48.85	48.083	49.439
Result	Mean Subst	Mean Subst	Mean Subst	Mean Subst	Mean Subst
Reference sample size					
Method	6	7	8	9	10
	6	7	8	9	10
<i>Chimp Test Sample</i>					
	$5.090 \times 10^{-10}$	$1.374 \times 10^{-9}$	$3.665 \times 10^{-10}$	$7.339 \times 10^{-12}$	$9.013 \times 10^{-13}$
	-7.676	-7.554	-7.911	-9.074	-9.659
	50.295	45.898	46.61	46.523	47.334
Result	TPS	TPS	TPS	TPS	TPS
<i>Gorilla Test Sample</i>					
	0.557	0.152	0.219	0.434	0.296
	0.592	1.455	1.243	0.789	1.056
	49.541	51.207	50.751	51.234	49.6
Result	—	—	—	—	—
<i>Human Test Sample</i>					
	$8.451 \times 10^{-10}$	$5.566 \times 10^{-4}$	$4.178 \times 10^{-6}$	$3.830 \times 10^{-7}$	$3.053 \times 10^{-6}$
	7.542	3.693	5.174	5.884	5.251
	50.11	49.098	49.403	47.733	50.352
Result	Mean Subst	Mean Subst	Mean Subst	Mean Subst	Mean Subst

Table 19: Two sided t-test results between the thin plate spline and mean substitution techniques. A chimp reference sample is used. These are the t-test results for analysis IV, section 6.1.4, which examines the effect of reference sample sizes; this table is specifically for small reference samples. The columns give how many individuals there are in the reference sample. For each t-test, if there is a significant difference ( $P < 0.05$ ) then the technique with the smaller mean residual is indicated.

<b>Gorilla Reference Sample</b>					
<b>Method</b>	<b>Reference sample size</b>				
	1	2	3	4	5
<i>Chimp Test Sample</i>					
<i>P</i>	0.006	0.082	0.177	0.208	$1.662 \times 10^{-4}$
<i>t</i>	-2.917	-1.774	-1.374	-1.277	-4.08
<i>df</i>	42.692	50.306	42.544	48.061	48.827
Result	TPS	—	—	—	TPS
<i>Gorilla Test Sample</i>					
<i>P</i>	$1.717 \times 10^{-5}$	$1.548 \times 10^{-8}$	$7.412 \times 10^{-7}$	$3.115 \times 10^{-5}$	$7.076 \times 10^{-4}$
<i>t</i>	-4.734	-6.646	-5.632	-4.558	-3.605
<i>df</i>	52.325	53.987	51.718	52.614	51.058
Result	TPS	TPS	TPS	TPS	TPS
<i>Human Test Sample</i>					
<i>P</i>	0.795	0.001	0.951	0.417	0.018
<i>t</i>	-0.261	3.455	-0.062	-0.819	-2.431
<i>df</i>	53.998	53.921	52.509	53.939	53.939
Result	—	Mean Subst	—	—	TPS
<b>Method</b>	<b>Reference sample size</b>				
	6	7	8	9	10
<i>Chimp Test Sample</i>					
	0.841	0.004	0.011	0.054	0.009
	0.202	-3.084	-2.665	-1.979	-2.706
	44.875	43.857	46.925	47.821	48.028
Result	—	TPS	TPS	—	TPS
<i>Gorilla Test Sample</i>					
	$7.699 \times 10^{-5}$	$3.222 \times 10^{-5}$	$3.918 \times 10^{-7}$	$4.402 \times 10^{-4}$	$1.342 \times 10^{-6}$
	-4.314	-4.556	-5.775	-3.766	-5.455
	49.279	51.647	53.935	49.731	52.536
Result	TPS	TPS	TPS	TPS	TPS
<i>Human Test Sample</i>					
	0.989	0.277	0.032	0.202	0.577
	0.014	-1.098	-2.209	-1.291	-0.561
	53.761	53.858	52.984	53.724	53.51
Result	—	—	TPS	—	—

Table 20: Two sided *t*-test results between the thin plate spline and mean substitution techniques. A gorilla reference sample is used. These are the *t*-test results for analysis IV, section 6.1.4, which examines the effect of reference sample sizes; this table is specifically for small reference samples. The columns give how many individuals there are in the reference sample. For each *t*-test, if there is a significant difference ( $P < 0.05$ ) then the technique with the smaller mean residual is indicated.

<b>Human Reference Sample</b>					
<b>Method</b>	<b>Reference sample size</b>				
	1	2	3	4	5
<i>Chimp Test Sample</i>					
<i>P</i>	0.023	0.001	0.003	$3.333 \times 10^{-6}$	$1.976 \times 10^{-5}$
<i>t</i>	2.346	3.461	3.164	5.238	4.748
<i>df</i>	50.646	51.257	50.005	49.45	46.787
Result	mean subst	mean subst	mean subst	mean subst	mean subst
<i>Gorilla Test Sample</i>					
<i>P</i>	0.495	0.036	0.570	0.065	0.331
<i>t</i>	0.688	2.15	0.572	1.882	0.981
<i>df</i>	51.484	51.152	52.204	52.424	51.354
Result	—	mean subst	—	—	—
<i>Human Test Sample</i>					
<i>P</i>	$4.364 \times 10^{-7}$	$3.995 \times 10^{-6}$	$2.737 \times 10^{-7}$	$1.492 \times 10^{-9}$	$2.738 \times 10^{-8}$
<i>t</i>	-5.788	-5.239	-6.082	-7.68	-6.938
<i>df</i>	51.151	45.672	43.169	42.344	38.627
Result	tps	tps	tps	tps	tps
<b>Method</b>	<b>Reference sample size</b>				
	6	7	8	9	10
<i>Chimp Test Sample</i>					
	0.888	0.392	$7.605 \times 10^{-4}$	0.030	0.024
	0.141	0.863	3.591	2.231	2.336
	51.008	50.945	49.141	49.539	49.428
Result	—	—	mean subst	mean subst	mean subst
<i>Gorilla Test Sample</i>					
	0.722	0.927	0.421	0.615	0.217
	-0.358	0.092	0.811	0.506	1.251
	52.234	51.735	51.829	50.849	50.724
Result	—	—	—	—	—
<i>Human Test Sample</i>					
	$4.407 \times 10^{-8}$	$8.233 \times 10^{-9}$	$8.696 \times 10^{-8}$	$2.053 \times 10^{-9}$	$5.464 \times 10^{-10}$
	-6.586	-6.896	-6.462	-7.517	-7.795
	44.499	50.626	41.895	43.854	46.591
Result	tps	tps	tps	tps	tps

Table 21: Two sided *t*-test results between the thin plate spline and mean substitution techniques. A human reference sample is used. These are the *t*-test results for analysis IV, section 6.1.4, which examines the effect of reference sample sizes; this table is specifically for small reference samples. The columns give how many individuals there are in the reference sample. For each *t*-test, if there is a significant difference ( $P < 0.05$ ) then the technique with the smaller mean residual is indicated.

Method	Reference Species		
	Chimp	Gorilla	Human
<i>Chimp Test Sample</i>			
Mean Substitution	0.529	0.667	0.845
Thin Plate Splines	0.371	0.621	0.898
Regression	—	—	0.530
<i>Gorilla Test Sample</i>			
Mean Substitution	0.970	0.792	1.408
Thin Plate Splines	1.001	0.598	1.438
Regression	—	—	0.905
<i>Human Test Sample</i>			
Mean Substitution	0.844	0.989	0.519
Thin Plate Splines	0.969	0.985	0.369
Regression	—	—	0.190

Table 22: The human reference column displays the mean residuals obtained using a reference sample of 600 individuals in figure 18. For comparison, the chimp and gorilla test samples used in the analysis are corrected using chimp ( $n = 29$ ) and gorilla ( $n = 64$ ) reference samples. Corrections via the regression-based method were not performed with these small reference samples.

	Reference Species		
	Chimp	Gorilla	Human
<i>Chimp Test Sample</i>			
<i>P</i>	$3.362 \times 10^{-12}$	0.028	0.033
<i>t</i>	9.213	2.265	-2.191
<i>df</i>	48.133	45.854	49.776
Result	<b>TPS</b>	<b>TPS</b>	<b>Mean Subst</b>
<i>Gorilla Test Sample</i>			
<i>P</i>	0.511	$9.250 \times 10^{-6}$	0.595
<i>t</i>	-0.662	4.920	-0.535
<i>df</i>	51.324	51.558	51.651
Result	—	<b>TPS</b>	—
<i>Human Test Sample</i>			
<i>P</i>	$1.557 \times 10^{-5}$	0.865	$2.239 \times 10^{-9}$
<i>t</i>	-4.793	0.170	7.424
<i>df</i>	49.149	53.671	45.552
Result	<b>Mean Subst</b>	<b>TPS</b>	<b>TPS</b>

Table 23: Two sided *t*-test results between the means obtained by the thin plate spline and mean substitution techniques (reported in table 22). The technique with the lower mean residual is reported if a two sided *t*-test indicates a significant difference between the means ( $P < 0.05$ ).

Method		Reference Species	
		Chimp	Gorilla
<i>Chimp Test Sample</i>			
vs. MS	<i>P</i>	0.937	$1.998 \times 10^{-10}$
	<i>t</i>	-0.080	7.870
	<i>df</i>	49.342	52.171
	Result	—	regression
vs. TPS	<i>P</i>	$5.056 \times 10^{-11}$	$1.446 \times 10^{-4}$
	<i>t</i>	-8.184	4.110
	<i>df</i>	53.883	50.72
	Result	tps	regression
<i>Gorilla Test Sample</i>			
vs. MS	<i>P</i>	0.097	0.002
	<i>t</i>	1.690	-3.139
	<i>df</i>	53.351	53.911
	Result	—	mean subst
vs. TPS	<i>P</i>	0.036	$4.741 \times 10^{-10}$
	<i>t</i>	2.156	-7.629
	<i>df</i>	48.688	52.333
	Result	regression	tps
<i>Human Test Sample</i>			
vs. MS	<i>P</i>	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$
	<i>t</i>	35.554	40.627
	<i>df</i>	47.068	44.516
	Result	regression	regression
vs. TPS	<i>P</i>	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$
	<i>t</i>	33.141	42.763
	<i>df</i>	38.907	46.629
	Result	regression	regression

Table 24: Two sided *t*-test results between the means obtained by the thin plate spline and mean substitution techniques (reported in table 22), and the means obtained with the regression technique using the human reference sample. Results are divided by the species of the test sample, the species of the reference sample used by the correction technique being tested against (MS being mean substitution, tps being thin plate splines). The technique with the lowest mean residual is reported if there are significant differences between the means ( $P < 0.05$ ).

**A.3 Tables For Analysis V: Cumulative Errors**

Method	Number of missing landmarks									
	1	2	3	4	5	6	7	8	9	10
<i>Chimp Test sample</i>										
Mean Subst	0.849	0.843	0.927	0.899	0.904	0.968	1.024	1.311	1.371	3.367
TPS	0.718	0.931	1.043	1.213	1.527	1.621	1.798	2.384	3.194	3.978
Regression	0.299	0.719	0.633	0.917	1.001	0.777	0.927	1.137	1.22	1.387
<i>Gorilla Test sample</i>										
Mean Subst	2.019	1.998	1.945	1.891	2.008	2.142	2.043	2.240	3.494	4.994
TPS	1.558	1.560	1.685	1.808	2.102	2.235	2.623	3.430	4.143	5.902
Regression	0.721	0.783	1.087	1.499	1.126	1.500	1.495	1.786	1.808	2.069
<i>Human Test sample</i>										
Mean Subst	0.430	0.415	0.425	0.443	0.478	0.454	0.476	0.490	0.612	2.795
TPS	0.443	0.620	0.796	0.992	1.368	1.575	1.844	2.427	2.497	4.080
Regression	0.188	0.219	0.219	0.240	0.232	0.276	0.294	0.329	0.343	0.405

Table 25: This table shows the mean residuals obtained by each technique when simultaneously estimating multiple missing landmarks. The columns give the number of missing landmarks, the rows show the techniques, with these being split into the three separate test samples.

Method	Number of missing landmarks									
	1	2	3	4	5	6	7	8	9	10
<b>Chimp Reference Sample</b>										
<i>Chimp Test Sample</i>										
Mean Subst	0.523	0.545	0.528	0.543	0.528	0.569	0.584	0.818	0.864	2.454
TPS	0.408	0.655	0.859	1.019	1.323	1.579	1.823	2.318	3.071	3.769
<i>Gorilla Test Sample</i>										
Mean Subst	1.478	1.564	1.423	1.485	1.487	1.594	1.768	1.722	2.171	3.597
TPS	1.216	1.111	1.414	1.427	1.826	2.168	2.610	3.119	3.607	5.361
<i>Human Test Sample</i>										
Mean Subst	1.087	0.917	0.933	1.062	0.987	1.068	1.022	1.371	1.740	3.332
TPS	0.945	1.109	1.204	1.288	1.565	1.709	2.058	2.502	3.130	4.307
<b>Gorilla Reference Sample</b>										
<i>Chimp Test Sample</i>										
Mean Subst	2.236	2.480	2.464	2.478	2.436	2.435	2.408	2.631	3.180	4.651
TPS	0.708	0.803	0.918	1.024	1.517	1.669	1.803	2.302	2.537	4.250
<i>Gorilla Test Sample</i>										
Mean Subst	1.306	1.237	1.250	1.259	1.269	1.314	1.327	1.496	1.885	5.428
TPS	0.814	0.984	1.167	1.446	1.716	2.021	2.820	3.035	3.483	5.541
<i>Human Test Sample</i>										
Mean Subst	2.578	2.682	2.690	2.745	2.691	2.743	2.801	2.814	3.360	4.198
TPS	1.221	1.286	1.474	1.570	1.770	1.785	2.210	2.321	3.216	4.256
<b>Human Reference Sample</b>										
<i>Chimp Test Sample</i>										
Mean Subst	1.093	0.891	0.954	0.983	0.975	0.975	1.094	1.166	2.010	3.085
TPS	0.897	0.982	1.151	1.235	1.440	1.648	1.975	2.379	3.031	4.395
<i>Gorilla Test Sample</i>										
Mean Subst	2.149	2.204	2.149	2.154	2.241	2.183	2.195	2.712	3.298	4.457
TPS	1.444	1.545	1.666	1.991	2.037	2.394	2.939	3.183	3.960	5.464
<i>Human Test Sample</i>										
Mean Subst	0.631	0.532	0.569	0.596	0.558	0.568	0.604	0.638	1.499	2.967
TPS	0.542	0.676	0.903	0.993	1.301	1.543	1.799	2.379	2.656	3.810

Table 26: This table shows the mean residuals obtained by the mean substitution and thin plate spline techniques when simultaneously estimating multiple missing landmarks using only a single reference individual. The columns give the number of missing landmarks, the rows show the techniques. The rows are divided by the species of both the reference sample and test sample.

No. Missing Landmarks	Test Species		
	Chimp	Gorilla	Human
1	$P = 0.328$ $t = 0.987$ $df = 47.361$ —	$P = 0.117$ $t = 1.596$ $df = 45.405$ —	$P = 0.815$ $t = -0.235$ $df = 53.325$ —
2	$P = 0.373$ $t = -0.899$ $df = 52.481$ —	$P = 0.048$ $t = 2.025$ $df = 44.552$ tps	$P = 0.001$ $t = -3.473$ $df = 42.618$ mean subst
3	$P = 0.241$ $t = -1.187$ $df = 49.952$ mean subst	$P = 0.117$ $t = 1.593$ $df = 53.482$ tps	$P = 5.319 \times 10^{-7}$ $t = -6.045$ $df = 37.253$ mean subst
4	$P = 1.669 \times 10^{-4}$ $t = -4.069$ $df = 50.314$ mean subst	$P = 0.602$ $t = 0.525$ $df = 53.779$ —	$P = 4.218 \times 10^{-10}$ $t = -8.260$ $df = 39.064$ mean subst
5	$P = 1.689 \times 10^{-5}$ $t = -5.009$ $df = 33.849$ mean subst	$P = 0.575$ $t = -0.564$ $df = 52.604$ —	$P = 1.947 \times 10^{-11}$ $t = -9.997$ $df = 32.474$ mean subst
6	$P = 5.754 \times 10^{-9}$ $t = -7.424$ $df = 38.887$ mean subst	$P = 0.652$ $t = -0.454$ $df = 40.194$ —	$P = 9.747 \times 10^{-16}$ $t = -14.692$ $df = 31.85$ mean subst
7	$P = 9.461 \times 10^{-9}$ $t = -7.359$ $df = 37.016$ mean subst	$P = 0.001$ $t = -3.397$ $df = 47.991$ mean subst	$P = < 2.2 \times 10^{-16}$ $t = -15.769$ $df = 34.246$ mean subst
8	$P = 2.635 \times 10^{-4}$ $t = -4.010$ $df = 39.285$ mean subst	$P = 3.247 \times 10^{-4}$ $t = -3.841$ $df = 53.969$ mean subst	$P = < 2.2 \times 10^{-16}$ $t = -21.538$ $df = 36.240$ mean subst
9	$P = 5.673 \times 10^{-9}$ $t = -7.126$ $df = 46.319$ mean subst	$P = 0.186$ $t = -1.347$ $df = 39.701$ —	$P = < 2.2 \times 10^{-16}$ $t = -13.389$ $df = 53.288$ mean subst
10	$P = 0.163$ $t = -1.420$ $df = 42.909$ —	$P = 0.090$ $t = -1.727$ $df = 51.997$ —	$P = 0.020$ $t = -2.433$ $df = 35.805$ mean subst

Table 27: Two sided  $t$ -test results between the mean substitution and thin plate spline techniques. These are  $t$ -test results for analysis V, section 6.1.5. The first column gives how many landmarks are being simultaneously corrected. For each  $t$ -test, if there is a significant difference ( $P < 0.05$ ) then the technique with the smaller mean residual is indicated.

No. Missing Landmarks	Test Species		
	Chimp	Gorilla	Human
1	$P = 6.927 \times 10^{-08}$ $t = 6.576$ $df = 40.523$ regression	$P = 2.111 \times 10^{-05}$ $t = 4.882$ $df = 36.350$ regression	$P = 1.157 \times 10^{-05}$ $t = 4.911$ $df = 46.569$ regression
2	$P = 0.345$ $t = 0.956$ $df = 42.499$ —	$P = 5.564 \times 10^{-06}$ $t = 5.059$ $df = 52.217$ regression	$P = 1.452 \times 10^{-06}$ $t = 5.480$ $df = 49.129$ regression
3	$P = 0.012$ $t = 2.632$ $df = 44.900$ regression	$P = 8.023 \times 10^{-06}$ $t = 4.956$ $df = 51.992$ regression	$P = 4.411 \times 10^{-08}$ $t = 6.407$ $df = 51.822$ regression
4	$P = 0.870$ $t = -0.164$ $df = 38.264$ —	$P = 0.042$ $t = 2.087$ $df = 48.406$ regression	$P = 7.550 \times 10^{-07}$ $t = 5.666$ $df = 49.154$ regression
5	$P = 0.415$ $t = -0.823$ $df = 34.719$ mean subst	$P = 1.352 \times 10^{-06}$ $t = 5.443$ $df = 53.319$ regression	$P = 7.368 \times 10^{-10}$ $t = 7.844$ $df = 43.437$ regression
6	$P = 0.042$ $t = 2.104$ $df = 38.056$ regression	$P = 0.006$ $t = 2.886$ $df = 48.652$ regression	$P = 5.217 \times 10^{-08}$ $t = 6.38$ $df = 50.937$ regression
7	$P = 0.268$ $t = 1.122$ $df = 42.501$ —	$P = 0.001$ $t = 3.389$ $df = 50.16$ regression	$P = 5.195 \times 10^{-06}$ $t = 5.183$ $df = 44.235$ regression
8	$P = 0.491$ $t = 0.696$ $df = 30.847$ regression	$P = 0.076$ $t = 1.824$ $df = 39.987$ —	$P = 2.026 \times 10^{-4}$ $t = 4.079$ $df = 41.156$ regression
9	$P = 0.326$ $t = 0.995$ $df = 36.833$ —	$P = 0.326$ $t = 0.995$ $df = 36.833$ —	$P = 0.018$ $t = 2.521$ $df = 28.569$ regression
10	$P = 1.403 \times 10^{-05}$ $t = 5.252$ $df = 27.922$ regression	$P = 8.641 \times 10^{-08}$ $t = 6.919$ $df = 31.4$ regression	$P = 4.118 \times 10^{-05}$ $t = 4.882$ $df = 27.179$ regression

Table 28: Two sided  $t$ -test results between the mean substitution and the regression-based techniques. These are  $t$ -test results for analysis V, section 6.1.5. The first column gives how many landmarks are being simultaneously corrected. For each  $t$ -test, if there is a significant difference ( $P < 0.05$ ) then the technique with the smaller mean residual is indicated.

No. Missing Landmarks	Test Species		
	Chimp	Gorilla	Human
1	$P = 0.001$ $t = 3.592$ $df = 33.502$ regression	$P = 4.201 \times 10^{-5}$ $t = 4.515$ $df = 47.323$ regression	$P = 9.843 \times 10^{-7}$ $t = 5.585$ $df = 49.531$ regression
2	$P = 0.124$ $t = 1.564$ $df = 47.058$ —	$P = 1.600 \times 10^{-4}$ $t = 4.087$ $df = 49.504$ regression	$P = 1.734 \times 10^{-8}$ $t = 7.227$ $df = 35.723$ regression
3	$P = 0.002$ $t = 3.327$ $df = 52.122$ regression	$P = 0.002$ $t = 3.311$ $df = 53.465$ regression	$P = 2.806 \times 10^{-11}$ $t = 9.671$ $df = 33.911$ regression
4	$P = 0.015$ $t = 2.526$ $df = 44.958$ regression	$P = 0.114$ $t = 1.61$ $df = 50.023$ —	$P = 1.483 \times 10^{-13}$ $t = 11.868$ $df = 33.549$ regression
5	$P = 0.001$ $t = 3.275$ $df = 53.793$ regression	$P = 8.549 \times 10^{-7}$ $t = 5.562$ $df = 53.861$ regression	$P = 9.728 \times 10^{-14}$ $t = 13.174$ $df = 28.875$ regression
6	$P = 1.000 \times 10^{-9}$ $t = 7.381$ $df = 53.914$ regression	$P = 2.682 \times 10^{-5}$ $t = 4.633$ $df = 49.1$ regression	$P = 3.824 \times 10^{-17}$ $t = 17.307$ $df = 29.955$ regression
7	$P = 3.448 \times 10^{-9}$ $t = 7.126$ $df = 50.994$ regression	$P = 2.430 \times 10^{-7}$ $t = 5.91$ $df = 53.608$ regression	$P = 6.793 \times 10^{-18}$ $t = 18.595$ $df = 29.656$ regression
8	$P = 1.044 \times 10^{-11}$ $t = 9.282$ $df = 41.759$ regression	$P = 4.023 \times 10^{-8}$ $t = 6.742$ $df = 40.561$ regression	$P = 2.400 \times 10^{-21}$ $t = 24.692$ $df = 29.691$ regression
9	$P = 4.992 \times 10^{-10}$ $t = 8.855$ $df = 31.262$ regression	$P = 5.246 \times 10^{-12}$ $t = 9.784$ $df = 38.628$ regression	$P = < 2.2 \times 10^{-16}$ $t = 22.597$ $df = 28.978$ regression
10	$P = 8.527 \times 10^{-13}$ $t = 11.837$ $df = 29.823$ regression	$P = 1.626 \times 10^{-12}$ $t = 10.862$ $df = 33.497$ regression	$P = < 2.2 \times 10^{-16}$ $t = 18.185$ $df = 28.065$ regression

Table 29: Two sided  $t$ -test results between the thin plate splines and the regression-based techniques. These are  $t$ -test results for analysis V, section 6.1.5. The first column gives how many landmarks are being simultaneously corrected. For each  $t$ -test, if there is a significant difference ( $P < 0.05$ ) then the technique with the smaller mean residual is indicated.

<b>Chimp Reference Sample</b>					
	<b>Number of missing landmarks</b>				
	1	2	3	4	5
<i>Chimp Test Sample</i>					
<i>P</i>	0.085	0.139	$2.489 \times 10^{-5}$	$1.131 \times 10^{-6}$	$2.247 \times 10^{-8}$
<i>t</i>	-1.749	1.505	4.828	5.873	7.364
<i>df</i>	53.473	46.603	36.324	35.047	31.997
Result	—	—	Mean Subst	Mean Subst	Mean Subst
<i>Gorilla Test Sample</i>					
<i>P</i>	0.238	0.012	0.950	0.679	0.052
<i>t</i>	-1.196	-2.624	-0.063	-0.416	1.991
<i>df</i>	46.219	46.775	49.87	51.416	53.058
Result	—	TPS	—	—	Mean Subst
<i>Human Test Sample</i>					
<i>P</i>	0.349	0.046	0.001	0.014	$2.046 \times 10^{-8}$
<i>t</i>	-0.946	2.049	3.433	2.554	6.969
<i>df</i>	45.008	51.39	53.667	53.457	40.093
Result	—	Mean Subst	Mean Subst	Mean Subst	Mean Subst
<b>Method</b>	<b>Number of missing landmarks</b>				
	6	7	8	9	10
<i>Chimp Test Sample</i>					
<i>P</i>	$1.399 \times 10^{-13}$	$5.728 \times 10^{-13}$	$9.188 \times 10^{-10}$	$2.771 \times 10^{-13}$	0.008
<i>t</i>	11.51	11.804	7.78	9.661	2.783
<i>df</i>	35.761	30.832	43.383	52.995	38.067
Result	Mean Subst	Mean Subst	Mean Subst	Mean Subst	Mean Subst
<i>Gorilla Test Sample</i>					
<i>P</i>	$1.678 \times 10^{-4}$	$8.173 \times 10^{-4}$	$2.848 \times 10^{-8}$	$8.171 \times 10^{-5}$	$5.857 \times 10^{-4}$
<i>t</i>	4.092	3.546	6.6	4.295	3.7
<i>df</i>	46.731	54	48.649	49.278	44.992
Result	Mean Subst	Mean Subst	Mean Subst	Mean Subst	Mean Subst
<i>Human Test Sample</i>					
<i>P</i>	$9.985 \times 10^{-9}$	$2.753 \times 10^{-11}$	$4.738 \times 10^{-5}$	$1.098 \times 10^{-4}$	0.060
<i>t</i>	6.855	9.32	4.52	4.175	1.941
<i>df</i>	50.141	37.352	43.32	53.67	36.458
Result	Mean Subst	Mean Subst	Mean Subst	Mean Subst	Mean Subst

Table 30: Two sided *t*-test results between the thin plate spline and the mean substitution techniques. A single chimp individual is used as the reference sample. These are *t*-test results for analysis V, section 6.1.5. The columns gives how many landmarks are being simultaneously corrected. For each *t*-test, if there is a significant difference ( $P < 0.05$ ) then the technique with the smaller mean residual is indicated.

<b>Gorilla Reference Sample</b>					
	<b>Number of missing landmarks</b>				
	1	2	3	4	5
<i>Chimp Test Sample</i>					
<i>P</i>	$3.704 \times 10^{-11}$	$1.597 \times 10^{-15}$	$1.306 \times 10^{-20}$	$6.75 \times 10^{-18}$	$1.829 \times 10^{-10}$
<i>t</i>	-9.438	-12.971	-15.122	-13.486	-7.886
<i>df</i>	35.059	37.95	51.648	47.723	52.423
Result	TPS	TPS	TPS	TPS	TPS
<i>Gorilla Test Sample</i>					
<i>P</i>	0.003	0.024	0.528	0.134	$1.891 \times 10^{-4}$
<i>t</i>	-3.126	-2.331	-0.636	1.52	4.048
<i>df</i>	49.623	53.507	51.386	51.999	47.608
Result	TPS	TPS	—	—	Mean Subst
<i>Human Test Sample</i>					
<i>P</i>	$8.510 \times 10^{-7}$	$3.995 \times 10^{-10}$	$7.305 \times 10^{-12}$	$2.137 \times 10^{-14}$	$1.976 \times 10^{-11}$
<i>t</i>	-5.802	-8.313	-9.504	-11.074	-8.436
<i>df</i>	40.645	38.528	40.456	44.669	53.947
Result	TPS	TPS	TPS	TPS	TPS
	<b>Number of missing landmarks</b>				
	6	7	8	9	10
<i>Chimp Test Sample</i>					
<i>P</i>	$3.436 \times 10^{-8}$	$3.222 \times 10^{-8}$	0.076	0.043	0.324
<i>t</i>	-6.584	-6.549	-1.808	-2.103	-0.998
<i>df</i>	47.199	49.333	51.437	33.635	37.998
Result	TPS	TPS	—	TPS	—
<i>Gorilla Test Sample</i>					
<i>P</i>	$4.806 \times 10^{-7}$	$5.803 \times 10^{-12}$	$2.138 \times 10^{-7}$	$5.787 \times 10^{-6}$	0.869
<i>t</i>	5.818	10.05	5.944	5.132	0.166
<i>df</i>	47.805	36.179	53.63	45.468	42.294
Result	Mean Subst	Mean Subst	Mean Subst	Mean Subst	—
<i>Human Test Sample</i>					
<i>P</i>	$5.207 \times 10^{-15}$	$2.128 \times 10^{-6}$	0.002	0.628	0.873
<i>t</i>	-11.047	-5.311	-3.197	-0.487	0.161
<i>df</i>	49.902	53.794	48.063	47.409	44.136
Result	TPS	TPS	TPS	—	—

Table 31: Two sided *t*-test results between the thin plate spline and the mean substitution techniques. A single gorilla individual is used as the reference sample. These are *t*-test results for analysis V, section 6.1.5. The columns gives how many landmarks are being simultaneously corrected. For each *t*-test, if there is a significant difference ( $P < 0.05$ ) then the technique with the smaller mean residual is indicated.

<b>Human Reference Sample</b>					
	<b>Number of missing landmarks</b>				
	1	2	3	4	5
<i>Chimp Test Sample</i>					
<i>P</i>	0.193	0.386	0.015	$3.810 \times 10^{-4}$	$6.956 \times 10^{-6}$
<i>t</i>	-1.319	0.873	2.516	3.795	5.2
<i>df</i>	50.319	53.921	49.283	52.999	38.303
Result	—	—	Mean Subst	Mean Subst	Mean Subst
<i>Gorilla Test Sample</i>					
<i>P</i>	0.009	$7.820 \times 10^{-4}$	0.012	0.351	0.241
<i>t</i>	-2.695	-3.611	-2.598	-0.942	-1.186
<i>df</i>	52.583	43.698	50.581	51.895	53.168
Result	TPS	TPS	TPS	—	—
<i>Human Test Sample</i>					
<i>P</i>	0.229	0.023	$3.122 \times 10^{-5}$	$4.417 \times 10^{-6}$	$1.054 \times 10^{-11}$
<i>t</i>	-1.217	2.337	4.627	5.224	9.615
<i>df</i>	49.953	53.964	45.284	44.725	37.822
Result	—	Mean Subst	Mean Subst	Mean Subst	Mean Subst
<b>Method</b>	<b>Number of missing landmarks</b>				
	6	7	8	9	10
<i>Chimp Test Sample</i>					
<i>P</i>	$1.616 \times 10^{-10}$	$5.963 \times 10^{-11}$	$1.661 \times 10^{-14}$	0.006	0.005
<i>t</i>	8.271	8.386	10.783	2.873	2.939
<i>df</i>	44.136	47.82	48.742	40.437	45.617
Result	Mean Subst	Mean Subst	Mean Subst	Mean Subst	Mean Subst
<i>Gorilla Test Sample</i>					
<i>P</i>	0.155	$1.866 \times 10^{-4}$	0.184	0.086	0.067
<i>t</i>	1.442	4.043	1.352	1.755	1.874
<i>df</i>	52.338	48.834	39.636	43.362	52.536
Result	—	Mean Subst	—	—	—
<i>Human Test Sample</i>					
<i>P</i>	$3.075 \times 10^{-15}$	$1.234 \times 10^{-12}$	$2.804 \times 10^{-14}$	0.005	0.089
<i>t</i>	12.557	10.74	12.77	2.969	1.742
<i>df</i>	38.775	35.076	32.812	34.859	40.158
Result	Mean Subst	Mean Subst	Mean Subst	Mean Subst	—

Table 32: Two sided *t*-test results between the thin plate spline and the mean substitution techniques. A single human individual is used as the reference sample. These are *t*-test results for analysis V, section 6.1.5. The columns gives how many landmarks are being simultaneously corrected. For each *t*-test, if there is a significant difference ( $P < 0.05$ ) then the technique with the smaller mean residual is indicated.

Mean Substitution Method					
	Number of missing landmarks				
	1	2	3	4	5
<i>Chimp Test Sample</i>					
<i>P</i>	0.087	0.612	0.716	0.229	0.216
<i>t</i>	1.75	0.51	0.365	1.218	1.251
<i>df</i>	45.333	53.147	51.175	53.691	53.462
Result	—	—	—	—	—
<i>Gorilla Test Sample</i>					
<i>P</i>	0.666	0.400	0.274	0.134	0.172
<i>t</i>	0.434	0.848	1.106	1.523	1.386
<i>df</i>	47.947	52.645	49.938	51.62	52.358
Result	—	—	—	—	—
<i>Human Test Sample</i>					
<i>P</i>	0.007	0.032	0.003	0.003	0.064
<i>t</i>	2.804	2.214	3.175	3.098	1.886
<i>df</i>	48.362	46.653	46.443	49.453	52.534
Result	multi	multi	multi	multi	—
Method	Number of missing landmarks				
	6	7	8	9	10
<i>Chimp Test Sample</i>					
<i>P</i>	0.907	0.342	0.564	0.072	0.597
<i>t</i>	0.118	0.959	-0.583	1.854	-0.532
<i>t</i>	53.573	48.729	30.951	37.144	53.997
Result	—	—	—	—	—
<i>Gorilla Test Sample</i>					
<i>P</i>	0.849	0.297	0.223	0.718	0.356
<i>t</i>	0.191	1.052	1.235	-0.364	-0.931
<i>df</i>	44.971	53.501	48.796	50.263	53.996
Result	—	—	—	—	—
<i>Human Test Sample</i>					
<i>P</i>	0.007	0.014	0.009	0.025	0.792
<i>t</i>	2.827	2.543	2.69	2.343	0.264
<i>df</i>	46.577	49.831	51.891	31.483	53.168
Result	multi	multi	multi	multi	—

Table 33: Two sided *t*-test results between the single-individual and multi-individual reference sample analyses of the mean substitution technique. These are *t*-test results for analysis V, section 6.1.5. Mean substitution is evaluated twice, once using only a single human reference individual and a second time with a sample of human individuals. The columns show how many landmarks are being simultaneously corrected. If there is a significant difference ( $P < 0.05$ ) in the mean residuals obtained with the two different samples, then the reference sample obtaining the smaller mean residual is indicated: **single** for the single reference individual, **multi** for the group.

<b>Thin Plate Spline Method</b>					
	<b>Number of missing landmarks</b>				
	1	2	3	4	5
<i>Chimp Test Sample</i>					
<i>P</i>	0.212	0.632	0.288	0.767	0.548
<i>t</i>	1.264	0.481	1.074	0.298	-0.605
<i>df</i>	51.936	53.998	51.737	48.772	48.081
Result	—	—	—	—	—
<i>Gorilla Test Sample</i>					
<i>P</i>	0.653	0.919	0.912	0.257	0.705
<i>t</i>	-0.451	-0.102	-0.111	1.146	-0.381
<i>df</i>	50.713	52.087	53.726	53.864	53.346
Result	—	—	—	—	—
<i>Human Test Sample</i>					
<i>P</i>	0.090	0.407	0.201	0.992	0.539
<i>t</i>	1.73	0.835	1.295	0.01	-0.618
<i>df</i>	52.408	52.374	53.599	53.639	52.191
Result	—	—	—	—	—
<b>Method</b>	<b>Number of missing landmarks</b>				
	6	7	8	9	10
<i>Chimp Test Sample</i>					
<i>P</i>	0.806	0.178	0.971	0.545	0.198
<i>t</i>	0.247	1.365	-0.037	-0.609	1.303
<i>df</i>	53.146	53.4	50.909	50.349	53.357
Result	—	—	—	—	—
<i>Gorilla Test Sample</i>					
<i>P</i>	0.233	0.130	0.357	0.524	0.365
<i>t</i>	1.205	1.539	-0.93	-0.642	-0.914
<i>df</i>	53.995	53.777	48.963	53.119	53.917
Result	—	—	—	—	—
<i>Human Test Sample</i>					
<i>P</i>	0.749	0.732	0.758	0.350	0.368
<i>t</i>	-0.322	-0.344	-0.31	0.944	-0.908
<i>df</i>	53.903	51.136	45.958	47.074	53.53
Result	—	—	—	—	—

Table 34: Two sided *t*-test results between the single-individual and multi-individual reference sample analyses of the thin plate spline method. These are *t*-test results for analysis V, section 6.1.5. Thin plate splines are evaluated twice, once using only a single human reference individual and a second time with a sample of human individuals. The columns show how many landmarks are being simultaneously corrected. If there is a significant difference ( $P < 0.05$ ) in the mean residuals obtained with the two different samples, then the reference sample obtaining the smaller mean residual is indicated: **single** for the single reference individual, **multi** for the group.

# Bibliography

- Rebecca Rogers Ackermann. Patterns of covariation in the hominoid craniofacial skeleton: implications for paleoanthropological models. *Journal of Human Evolution*, 43:167–187, 2002.
- Rebecca Rogers Ackermann. Ontogenetic integration in the hominoid face. *Journal of Human Evolution*, 2005.
- Rebecca Rogers Ackermann. *A quantitative assessment of variability in the australopithecine, human, chimpanzee, and gorilla face*. PhD thesis, Washington University, St Louis, 1998.
- Rebecca Rogers Ackermann and James M. Cheverud. Phenotypic covariance structure in tamarins (genus *Saguinus*): A comparison of variation patterns using matrix correlation and common principal component analysis. *American Journal of Physical Anthropology*, 111:489–501, 2000.
- Rebecca Rogers Ackermann and James M. Cheverud. Discerning evolutionary processes in patterns of tamarin (genus *Saguinus*) craniofacial variation. *American Journal of Physical Anthropology*, 117:260–271, 2002.
- Leslie Aiello, Bernard Wood, Cathy Key, and Christopher Wood. Laser scanning and palaeoanthropology: an example from olduvai gorge, tanzania. In E Strasser, J Fleagle, A Rosenberger, and H McHenry, editors, *Primate Locomotion: Recent Advances*. Plenum Press, New York, 1998.
- Nina Amenta and Marshall Bern. Surface reconstruction by voronoi filtering. In *SCG '98: Proceedings of the fourteenth annual symposium on Computational geometry*, pages 39–48, New York, NY, USA, 1998. ACM Press. ISBN 0-89791-973-4. doi: <http://doi.acm.org/10.1145/276884.276889>.
- Nina Amenta, Marshall Bern, and Manolis Kamvysselis. A new voronoi-based surface reconstruction algorithm. In *SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 415–421, New York, NY, USA, 1998. ACM Press. ISBN 0-89791-999-8. doi: <http://doi.acm.org/10.1145/280814.280947>.

- P. R. Andresen and M. Nielsen. Non-rigid registration by geometry-constrained diffusion. *Medical Image Analysis*, 5:81–88, 2001.
- B. Asfaw, T. White, O. Lovejoy, B. Latimer, S. Simpson, and G. Suwa. *Australopithecus garhi*: a new species of early hominid from ethiopia. *Science*, 284:629–635, 1999.
- K. B. Atkinson, editor. *Close Range Photogrammetry and Machine Vision*. Whittles Publishing, Scotland, UK, 1996.
- D. Bechmann. Space deformation models survey. *Computers and Graphics*, 18(4):571–586, 1994.
- Fausto Bernardini and Holly Rushmeier. The 3d model acquisition pipeline. *Computer Graphics Forum*, 21:149–172, 2002.
- P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
- Harry Blum. Biological shape and visual science (part i). *Journal of Theoretical Biology*, 38: 205–287, 1973.
- Jean-Daniel Boissonnat. Geometric structures for three-dimensional shape representation. *ACM Trans. Graph.*, 3(4):266–286, 1984. ISSN 0730-0301. doi: <http://doi.acm.org/10.1145/357346.357349>.
- Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989. doi: <http://dx.doi.org/10.1109/34.24792>.
- Fred. L. Bookstein. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, New York, 1991.
- Fred L. Bookstein. Size and shape spaces for landmark data in two dimensions (with discussion). *Statistical Science*, 1:181–242, 1986.
- Fred L. Bookstein. Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. *Medical Image Analysis*, 1(3):225–243, 1997.
- Fred L. Bookstein, Philipp Gunz, Philipp Mitteroecker, Hermann Prossinger, Katrin Schaefer, and Horst Seidler. Cranial integration in *Homo*: singular warps analysis of the midsagittal plane in ontogeny and evolution. *Journal of Human Evolution*, 44:167–187, 2003.
- Martin Brady, Kenneth Jung, HT Nguyem, and Thinh Nguyen. Two phase perspective ray casting for interactive volume navigation. *IEEE Visualisation*, 1997.

C. K. Brain. *The Hunters or the Hunted?: An Introduction to African Cave Taphonomy*. University of Chicago Press, Chicago, 1981.

Michel Brunet, Franck Guy, David Pilbeam, Hassane Taisso Mackaye, Andossa Likius, Djimdoumalbaye Ahounta, Alain Beauvilain, Cécile Blondel, Hervé Boucherns, Jean-Renaud Boisserie, Louis de Bonis, Yves Coppens, Jean Dejax, Christiane Denys, Philippe Duringer, Véra Eisenmann, Gongdibé Fanone, Pierre Fronty, Denis Geraads, Thomas Lehmann, Fabrice Lihoreau, Antoine Louchart, Adoum Mahamat, Gildas Merceron, Guy Mouchelin, Olga Otero, Pablo Pelaez Campomanes, Marcia Ponce de Leon, Jean-Claude Rage, Michel Sapanet, Mathieu Schuster, Jean Sudre, Pascal Tassy, Xavier Valentin, Patrick Vignaud, Laurent Viriot, Antoine Zazzo, and Christoph Zollikofer. A new hominid from the upper miocene of chad, central africa. *Nature*, 418:145–151, 2002.

P. Cignoni, C. Rocchini, and R. Scopigno. Metro: measuring error on simplified surfaces. *Computer Graphics Forum*, 17(2), 1998.

R. J. Clarke. A juvenile cranium and some adult teeth of early *Homo* from swartkrans, transvaal. *South African Journal of Science*, 73:46–49, 1977.

Glenn C. Conroy. *Reconstructing Human Origins : a Modern Synthesis*. W. W. Norton, New York, 1997.

Glenn C. Conroy and Michael W. Vannier. Dental development of the taung skull from computerized tomography. *Nature*, 329, 1987.

Glenn C. Conroy and Michael W. Vannier. The taung skull revisited: new evidence from high-resolution computed tomography. *South African Journal of Science*, 85, 1989.

Glenn C. Conroy, Gerhard W. Weber, Horst Seidler, Phillip V. Tobias, Alex Kane, and Barry Brunsden. Endocranial capacity in an early hominid cranium from sterkfontein, south africa. *Science*, 280, 1998.

Glenn C. Conroy, Gerhard W. Weber, Horst Seidler, and Phillip V. Tobias. Technical comments: Endocranial capacity of early hominids. *Science*, 283, 1999.

Glenn C. Conroy, Dean Falk, John Guyer, Gerhard W. Weber, Horst Seidler, and Wolfgang Recheis. Endocranial capacity in Sts 71 (*Australopithecus africanus*) by three-dimensional computed tomography. *The Anatomical Record*, 258:391–396, 2000.

Trevor F. Cox and Michael A. A. Cox. *Multidimensional Scaling*. Chapman & Hall, London, 1994.

Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 11–20. ACM Press, 1996. ISBN 0-89791-746-4. doi: <http://doi.acm.org/10.1145/237170.237191>.

Paul E. Debevec, Camillo J. Taylor, Jitendra Malik, Golan Levin, George Borshukov, and Yizhou Yu. Modeling and rendering of architecture with interactive photogrammetry and view-dependent texture mapping. *International Symposium on Circuits and Systems (ISCAS)*, 1998.

Tony DeRose, Hugues Hoppe, Tom Duchamp, John McDonald, and Werner Stuetzle. Fitting of surfaces to scattered data. In *SPIE*, volume 1830, pages 212–220, 1992.

W. Dieck. Das gebiss des diluvialen homo mousteriensis hauseri und seine rekonstruktion: Vortrag auf der jahresversammlung der schwedischen zahnärztlichen gesellschaft in goteborg, august 1923. *Odontologisk Tidskrift*, 3:196–209, 1923.

Pierre Drap and Luc Long. Towards a digital excavation data management system: the "grand ribaud f" estruscan deep-water wreck. In *Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage*, pages 17–26. ACM Press, 2001. ISBN 1-58113-447-9. doi: <http://doi.acm.org/10.1145/584993.584997>.

Ian L. Dryden and Kanti V. Mardia. *Statistical Shape Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd, Chichester, 1998.

E. Edelsbrunner, D. G. Kirkpatrick, and R Seidel. On the shape of points in the plane. *IEEE Transactions on Information Theory*, 29(4):551–559, 1983.

Herbert Edelsbrunner and Ernst P. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graph.*, 13(1):43–72, 1994. ISSN 0730-0301. doi: <http://doi.acm.org/10.1145/174462.156635>.

Gerald E. Farin. *Curves and Surfaces for Computer Aided Geometric Design*. Academic Press, Boston, 1992.

Donald E. Farrar and Robert R. Glauber. Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics*, 49(1):92 – 107, 1967.

Olivier Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, Cambridge, Mass, 1993.

James D. Foley, Andries van Dam, Steven K. Feiner, and John F. Hughes. *Computer Graphics: Principles and Practice*. Addison-Wesley Publishing Company, second in c edition, 1996.

Thomas Funkhouser, Patrick Min, Michael Kazhdan, Joyce Chen, Alex Halderman, David Dobkin, , and David Jacobs. A search engine for 3d models. *ACM Transactions on Graphics*, 22(1):83–105, 2003.

Thomas Funkhouser, Michael Kazhdan, Patrick Min, and Philip Shilane. Shape-based retrieval and analysis of 3d models. *Communications of the ACM*, 48(6):58–64, 2005.

Sarah F. Gibson and Brian Mirtich. A survey of deformable models in computer graphics. Technical Report TR-97-19, Mitsubishi Electric Research Laboratories, Cambridge, MA, November 1997.

R. Gonzalez and P. Wintz. *Digital Image Processing*. Addison-Wesley, Reading, Massachusetts, 1987.

Philipp Gunz. *Statistical & Geometric Reconstruction of Hominid Crania: Reconstructing Australopithecine Ontogeny*. PhD thesis, Universität Wien, Wien, September 2005.

Philipp Gunz, Philipp Mitteroecker, and Fred L. Bookstein. Semilandmarks in three dimensions. In D. E. Slice, editor, *Modern Morphometrics in Physical Anthropology*. Kluwer Academic / Plenum Publishers, 2005.

W. Härdle. *Applied Nonparametric Regression*. Cambridge University Press, 1990.

Paul S. Heckbert. Fundamentals of texture mapping and image warping. Master's thesis, University of California, 1989.

Hugues Hoppe, Tony DeRose, Tom Duchamp, John McDonald, and Werner Stuetzle. Surface reconstruction from unorganized points. In *SIGGRAPH '92: Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, pages 71–78, New York, NY, USA, 1992. ACM Press. ISBN 0-89791-479-1. doi: <http://doi.acm.org/10.1145/133994.134011>.

Hugues Hoppe, Tony DeRose, Tom Duchamp, John McDonald, and Werner Stuetzle. Mesh optimization. In *SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 19–26, New York, NY, USA, 1993. ACM Press. ISBN 0-89791-601-8. doi: <http://doi.acm.org/10.1145/166117.166119>.

Hugues Hoppe, Tony DeRose, Tom Duchamp, Mark Halstead, Hubert Jin, John McDonald, Jean Schweitzer, and Werner Stuetzle. Piecewise smooth surface reconstruction. In *SIGGRAPH '94: Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 295–302, New York, NY, USA, 1994. ACM Press. ISBN 0-89791-667-0. doi: <http://doi.acm.org/10.1145/192161.192233>.

- Berthold K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, 4:629–642, 1987.
- W. M. Hsu, J. F. Hughes, and H. Kaufman. Direct manipulation of free-form deformations. *Computer Graphics*, 26(2):177–184, 1992.
- Ch. Ioannidis and Maria Tsakiri. Laser scanning and photogrammetry for the documentation of a large statue — experiences in the combined use. In *Proc. CIPA XIXth Int. Symposium*, pages 517–523, 2003.
- I. T. Jolliffe. Discarding variables in a principal component analysis. i: Artificial data. *Applied Statistics*, 21(2):160 – 173, 1972.
- I. T. Jolliffe. Discarding variables in a principal component analysis. ii: Real data. *Applied Statistics*, 22(1):21 – 31, 1973.
- I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag, New York, 1986.
- Michel Kasser and Yves Egels. *Digital Photogrammetry*. Taylor & Francis, 11 New Fetter Lane, London EC4P 4EE, 2002.
- Susan M. Kidwell and Steven M. Holland. The quality of the fossil record: Implications for evolutionary analyses. *Annual Review of Ecology and Systematics*, 33:561–588, 2002.
- H. Klaatsch. Die fortschritte der lehre von der neandertalrasse. *Ergebnisse anatomischer Entwicklungsgeschichte*, 17:431–462, 1909.
- H. Klaatsch and O. Hauser. Homo mousteriensis hauseri: ein altdiluvialer skelettfund im department dordogne und seine zugehörigkeit zum neandertaltypus. *Arch Anthropol*, 7:287–297, 1908.
- Richard G. Klein. *The Human Career: Human Biological and Cultural Origins*. University of Chicago Press, Chicago, 2nd edition, 1999.
- Karl Kraus. *Photogrammetry*, volume 2. Ümmler, Bonn, 1997.
- Kent Leahy. Multicollinearity: When the solution is the problem. In Olivia Parr Rud, editor, *Data Mining Cookbook*. John Wiley & Sons, Inc., 2000.
- Marc Levoy and Turner Whitted. The use of points as a display primitive. Technical Report TR 85-022, The University of North Carolina at Chapel Hill, Computer Science Department, 1985.
- Mark Levoy. Efficient ray tracing of volume data. *ACM Transactions on Graphics*, 1990.

- Bruno Lévy, Sylvain Petitjean, Nicolas Ray, and Jérôme Maillot. Least squares conformal maps for automatic texture atlas generation. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 362–371, New York, NY, USA, 2002. ACM Press. ISBN 1-58113-521-1. doi: <http://doi.acm.org/10.1145/566570.566590>.
- William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 163–169, New York, NY, USA, 1987. ACM Press. ISBN 0-89791-227-6. doi: <http://doi.acm.org/10.1145/37401.37422>.
- R. L Lyman. *Vertebrate Taphonomy*. Cambridge University Press, New York, 1994.
- M. Martinón-Torres, M. Bastir, J. M Bermúdez de Castro, A. Gómez, S. Sarmiento, A. Muela, and J. L. Arsuaga. Hominin lower second premolar morphology: evolutionary inferences through geometric morphometric analysis. *Journal of Human Evolution*, 1(11), 2006.
- Marcia S. Ponce De León and Christoph P. E. Zollikofer. New evidence from le moustier 1: Computer-assisted reconstruction and morphometry of the skull. *The Anatomical Record*, 254: 474–489, 1999.
- Edward M. Mikhail, James S. Bethel, and J. Chris McGlone. *Introduction to Modern Photogrammetry*. John Wiley & Sons, Inc, United States of America, 2001.
- Tim Milliron, Robert J. Jensen, Ronen Barzel, and Adam Finkelstein. A framework for geometric warps and deformations. *ACM Trans. Graph.*, 21(1):20–51, 2002. ISSN 0730-0301. doi: <http://doi.acm.org/10.1145/504789.504791>.
- Paul Ning and Jules Bloomenthal. An evaluation of implicit surface tilers. *IEEE Computer Graphics and Applications*, 13(6):33–41, 1993.
- Paul O'Higgins. The study of morphological variation in the hominid fossil record: Biology, landmarks and geometry. *Journal of Anatomy*, 197:103–120, 2000.
- Norbert Pfeifer and Derek Lichti. Terrestrial laser scanning. *GIM International*, 18:50–53, December 2004.
- Hanspeter Pfister, Matthias Zwicker, Jeroen van Baar, and Markus Gross. Surfels: Surface elements as rendering primitives. In *SIGGRAPH 2000*, 2000.
- L. J. Pilgrim. History of photogrammetry in medicine. *Australas Phys Eng Sci Med*, 15:1–8, 1992.
- Pierre Poulin, Mathieu Ouimet, and Marie-Claude Frasson. Interactively modeling with photogrammetry. In *Proc. Eurographics Workshop on Rendering 98*, 1998.

- Vaughan Pratt. Direct least-squares fitting of algebraic surfaces. In *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer Graphics and Interactive Techniques*, volume 21, pages 145–152, 1987.
- F. P. Preparata and M. I. Shamos. *Computational Geometry — An Introduction*. Springer-Verlag, New York, 1985.
- Long Quan and Zhongdan Lan. Linear n-point camera pose estimation. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 21(8):774–780, August 1999.
- R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- Joan T. Richtsmeier, James M. Cheverud, and Subhash Lele. Advances in anthropological morphometrics. *Annual Review of Anthropology*, 21:283–305, 1992.
- Stefan Roettger, Martin Kraus, and Thomas Ertl. Hardware-accelerated volume and isosurface rendering based on cell-projection. *IEEE Visualisation*, 2000.
- Szymon Rusinkiewicz and Marc Levoy. Qsplat: a multiresolution point rendering system for large meshes. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 343–352, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co. ISBN 1-58113-208-5. doi: <http://doi.acm.org/10.1145/344779.344940>.
- Maryellen Ruvolo. Molecular phylogeny of the hominoids: inferences from multiple independent dna sequence data sets. *Molecular Biology and Evolution*, 14:248–265, 1997.
- C. Schuchardt. Die neue zusammensetzung des schadels vom homo moustieriensis hauseri. *Amtliche Berichte aus den Koniglichen Kunstsammlungen*, 34(10), 1912.
- T. Schulz and H. Ingensand. Terrestrial laser scanning — investigations and applications for high precision scanning. In *Proc. FIG Working Week*, 2004.
- T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. *Computer Graphics*, 20:151–169, 1986.
- S. Seeger and X. Labourey. Feature extraction and registration — an overview. In B. Girod, G. Greiner, and H. Niemann, editors, *Principles of 3D Image Analysis and Synthesis*. Kluwer Academic Publishers, London, 2000.
- Brigitte Senut, Martin Pickford, Dominique Gommery, Pierre Mein, Kiptalam Cheboi, and Yves Coppens. First hominid from the miocene (lukino format, kenya). *Human Palaeontology*, 2001.

- Christopher G. Small. *The Statistical Theory of Shape*. Springer Series in Statistics. Springer-Verlag, New York, 1996.
- Julian Loyd Smit. *Three Dimensional Measurement of Textured Surfaces using Digital Photogrammetric Techniques*. PhD thesis, University of Cape Town, 1997.
- Richard J. Smith. Species recognition in paleoanthropology: implications of small sample sizes. In D. Lieberman, R. Smith, and J. Kelley, editors, *Interpreting the Past: Essays on Human, Primate, and Mammal Evolution in Honor of David Pilbeam*, pages 207–219. Brill Academic Publishers, Boston, 2005.
- R. R. Sokal and F. J. Rohlf. *Biometry*. Freeman, New York, second edition, 1995.
- Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image Processing, Analysis, and Machine Vision*. Brooks/Cole Publishing Company, Pacific Grove, CA, 2nd edition, 1999.
- Fred Spoor, Nathan Jeffery, and Frans Zonneveld. Using diagnostic radiology in human evolutionary studies. *Journal of Anatomy*, 197:61–76, 2000.
- D. Terzopoulos. Multilevel computational processes for visual surface reconstruction. *Computer vision, graphics and image processing*, 24:52–96, 1983.
- Bill Triggs, Philip McLauchlan, Richard Hartley, and Andrew Fitzgibbon. Bundle adjustment – A modern synthesis. In W. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag, 2000.
- Greg Turk and Marc Levoy. Zippered polygon meshes from range images. In *SIGGRAPH '94: Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 311–318, New York, NY, USA, 1994. ACM Press. ISBN 0-89791-667-0. doi: <http://doi.acm.org/10.1145/192161.192241>.
- Patrick Vignaud, Philippe Duringer, Hassane Taïsso Mackaye, Andossa Likius, Cécile Blondel, Jean-Renaud Boisserie, Louis de Bonis, Véra Eisenmann, Marie-Ester Etienne, Denis Gerads, Franck Guy, Thomas Lehmann, Fabrice Lihoreau, Nieves Lopez-Martinez, Cécile Mourer-Chauviré, Olga Otero, Jean-Claude Rage, Mathieu Schuster, Laurent Viriot, Antoine Zazzo, and Michel Brunet. Geology and palaeontology of the upper miocene toros-menalla hominid locality, chad. *Nature*, 418:152–155, 2002.
- Gerhard W. Weber. Virtual anthropology (VA): a call for *glasnost* in paleoanthropology. *The Anatomical Record*, 265:193–201, 2001.

H. Weinert. *Der Schadel des eiszeitlichen Menschen von Le Moustier in neuer Zusammensetzung.* Springer, Berlin, 1925.

T. D. White, G. Suwa, and B. Asfaw. *Australopithecus ramidus*, a new species of hominid from aramis, ethiopia. *Nature*, 371:306–312, 1994.

Tim White. Early hominids — diversity or distortion? *Science*, 300, 2003.

Jane Wilhelms and Allen Van Gelder. Topological considerations in isosurface generation, extended abstract. *Computer Graphics, Special Issue on San Diego Workshop on Volume Visualisation*, 24(5):79–86, 1990.

Bernard Wood. Investigating human evolutionary history. *Journal of Anatomy*, 197:3–17, 2000.

Bernard Wood and Brian G. Richmond. Human evolution: taxonomy and paleobiology. *Journal of Anatomy*, 196:19–60, 2000.

Bernard Wood, Leslie Aiello, Christopher Wood, and Cathy Key. A technique for establishing the identity of ‘isolated’ fossil hominin limb bones. *Journal of Anatomy*, 193:61–72, 1998.

Christoph P. E. Zollikofer and Marcia S. Ponce de León. Tools for rapid prototyping in the biosciences. *IEEE Computer Graphics and Applications*, 15:48–55, 1995.

Christoph P. E. Zollikofer and Marcia S. Ponce de León. *Virtual Reconstruction: A Primer in Computer-Assisted Paleontology and Biomedicine*. John Wiley & Sons, Inc, Hoboken, New Jersey, 2005.

Christoph P. E. Zollikofer, Marcia S. Ponce de León, Robert D. Martin, and Peter Stucki. Neanderthal computer skulls. *Nature*, 375:283–285, 1995.

Christoph P. E. Zollikofer, Marcia S. Ponce de León, and Robert D. Martin. Computer-assisted paleoanthropology. *Evolutionary Anthropology*, 6:41–54, 1998.

Christoph P. E. Zollikofer, Marcia S. Ponce de León, F. Esteves, F. Tecelao Silva, and R. Pacheco Dias. The computer-assisted reconstruction of the skull. In J. Zilhão and E. Trinkaus, editors, *Portrait of the Artist as a Child. The Gravettian Human Skeleton from the Abrigo do Lagar Velho and its Archeological Context.*, pages 326–341. Instituto Português de Arqueologia, Lisbon, 2002a.

Christoph P. E. Zollikofer, Marcia S. Ponce de León, Bernard Vandermeersch, and Francois Leveque. Evidence for interpersonal violence in the St. Césaire Neanderthal. *PNAS*, 99(9):6444–6448, 2002b.

Christoph P. E. Zollikofer, Marcia S. Ponce de León, Daniel E. Lieberman, Franck Guy, David Pilbeam, Andossa Likius, Hassane T. Mackaye, Patrick Vignaud, and Michael Brunet. Virtual cranial reconstruction of *Sahelanthropus tchadensis*. *Nature*, 434:755 – 759, 7 April 2005.

Matthias Zwicker, Hanspeter Pfister, Jeroen van Baar, and Markus Gross. Surface splatting. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 371–378, New York, NY, USA, 2001. ACM Press. ISBN 1-58113-374-X. doi: <http://doi.acm.org/10.1145/383259.383300>.

And finally,

*fin*

“Go now — go!”

“I’ll just try to think, ‘Where the hell’s the whiskey?’”