http://researchspace.auckland.ac.nz

ResearchSpace@Auckland

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage. http://researchspace.auckland.ac.nz/feedback

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.
Development of an Automated System for Building a Large Population-based Statistical Model of Femur Morphology

Ju Zhang

Supervised by Professor Poul Nielsen and Dr Duane Malcolm

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Bioengineering, The University of Auckland 2013

Auckland Bioengineering Institute
The University of Auckland
New Zealand
2013
Abstract

Studying femur morphology on a large population of computed-tomography (CT) images requires automatic methods. This thesis presents a fully automatic CT-to-model pipeline that accurately segments and models femur morphology. The pipeline is composed of a training phase, where a statistical shape model is created, and a processing phase, which segments and models cortical bone geometry and cancellous bone mineral density (BMD) distribution. Development and testing of the pipeline was carried out on a set of 262 quantitative-CT images from the Victorian Institute of Forensic Medicine (VIFM).

In the training phase, corresponding regions on a training-set of 41 femoral surfaces were automatically partitioned and grouped using region-growing and mean-shift clustering. These regions were used to design a region-based quartic-Lagrange femur mesh, which was fitted region-by-region to manually segmented surfaces, to train the femur statistical shape model. Validation experiments showed that this region-based shape model was more accurate and correspondent than an equivalent non-regional model.

Cortical bone geometry was automatically extracted and modelled in the first step of the processing phase, using the shape mode above. Active shape modelling and cortical thickness mapping were adapted and combined to mesh the inner and outer cortical surfaces. Segmented meshes were accurate to 0.9 mm root-mean-square (RMS), and cortical thickness to 0.6 mm RMS. The method achieved a success rate of 83%.

Cancellous BMD images were automatically segmented and registered in the second step of the processing phase. BMD values from CT images were mapped to a reference volume by radial basis functions (RBFs), which interpolated the mapping between segmented and reference inner cortical surface meshes. Compared to conventional free-form deformation (FFD) registration, RBF registration followed by FFD led to a four-fold reduction in run time, a surface accuracy of 0.76 mm (versus 3.7 mm), and better alignment of anatomical features. Principal component analysis of registered images showed BMD variations in clinically relevant regions.

The development of the CT-to-model pipeline has enabled unsupervised data collection from VIFM CT scans for scientific and clinical studies of femur morphology. As a general framework, minor modifications of the pipeline will also allow unsupervised data collection for other bones and other image sets.
Publications

This body of research has been presented at the following local and international conferences:


Acknowledgements

First and foremost, I would like to thank my supervisors Professor Poul Nielsen and Dr Duane Malcolm. Their special blend of wisdom and pragmatism supported and guided me through this adventure.

I am very grateful for the help of Mr David Thomas, Professor John Clement, and Jacqui Hislop-Jambrich. Their knowledge of all things bone-related has been invaluable, and I look forward to our continued collaboration in the future. Special thanks goes to Jacqui for her encouragement and advice regarding the thesis. And of course, none of this work would have been possible without the generosity of the Victorian Institute of Forensic Medicine, for supporting Jacqui in accessing and collecting the femur CT images. I would also like to thank Dr Andrew Gee and Dr Graham Treece for their help with cortical thickness mapping, and their warm hospitality during my visits to Cambridge.

I would like to thank my colleagues at the Auckland Bioengineering Institute, the admin team, the IT staff, the Friday Drinks crew over the years. You guys make this place great.

Many thanks to all of my friends, inside and outside of the Institute, for the technical (and non-technical) discussions, correcting my typos, providing a couch on the other side of the world, or just smack talk around the lunch room. Special mention goes out to Simon, Oliver, Prasad, the ABI lunch group, Pat & Niv, and of course the proof-reading team (in no particular order: Aditi, Dharshini, Sophia, Michelle, and Ioannis). It was great to have you all along for the ride.

I want to express my deep gratitude for my family: Mum and Dad, thank you for 26 years of selfless sacrifice, patience, and understanding. I would not be where I am today without you. Alec, thanks for being my little bro. And finally, to my extended family, sorry for being so rare these four years, I'll come visit soon.

Funding from the University of Auckland Doctoral Scholarship and the ABI was greatly appreciated.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Publications</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>vii</td>
</tr>
<tr>
<td>Contents</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xx</td>
</tr>
<tr>
<td>Glossary of Acronyms</td>
<td>xxi</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivations</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Central Aims and Scope</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Novel Contributions</td>
<td>3</td>
</tr>
<tr>
<td>1.4 Thesis and Pipeline Overview</td>
<td>3</td>
</tr>
<tr>
<td><strong>2 Statistical Shape Analysis</strong></td>
<td>9</td>
</tr>
<tr>
<td>2.1 Shape Descriptions</td>
<td>10</td>
</tr>
<tr>
<td>2.1.1 Point Distribution Models (PDMs)</td>
<td>10</td>
</tr>
<tr>
<td>2.1.2 Piecewise Parametric Meshes (PPMs)</td>
<td>11</td>
</tr>
<tr>
<td>2.1.3 Medial Descriptors</td>
<td>12</td>
</tr>
<tr>
<td>2.1.4 Statistical Deformation Models</td>
<td>13</td>
</tr>
<tr>
<td>2.1.5 Global Descriptors</td>
<td>14</td>
</tr>
</tbody>
</table>
2.1.6 Summary ................................................. 15
2.2 Alignment ................................................. 16
2.3 Statistical Analysis ...................................... 17
  2.3.1 Dimension Reduction ................................. 17
  2.3.2 Correlation ........................................... 20
  2.3.3 Classification ....................................... 21
2.4 Shape Description Correspondence .................... 22
  2.4.1 Point-to-Point ....................................... 22
  2.4.2 Parameterisation .................................... 24
  2.4.3 Population-based .................................... 25
  2.4.4 Model Selection ..................................... 27
  2.4.5 Summary ............................................ 28
2.5 Final Remarks ............................................ 29
3 Femur Morphology ......................................... 31
  3.1 Femur Anatomy ......................................... 31
    3.1.1 External Features ................................ 32
    3.1.2 Cortical Bone ..................................... 37
    3.1.3 Cancellous Bone .................................. 38
  3.2 Bone Biology ........................................... 39
    3.2.1 Composition ....................................... 39
    3.2.2 Cellular Components ............................ 40
    3.2.3 Bone Remodelling ................................. 40
  3.3 Traditional Measurements ............................. 42
    3.3.1 Femoral Lengths ................................ 42
    3.3.2 Femoral Widths ................................... 43
    3.3.3 Axes and Angles .................................. 43
    3.3.4 Femoral Cortex .................................. 44
    3.3.5 Cancellous Bone .................................. 45
    3.3.6 Bone Mineral Density ............................ 45
  3.4 Natural Variations ..................................... 47
    3.4.1 Development ....................................... 47
## CONTENTS

3.4.2 Ageing ................................................. 48
3.4.3 Gender .............................................. 50
3.4.4 Ancestry ............................................ 51
3.4.5 Stature, Body Mass and Other Factors .................. 51
3.5 Pathologies ............................................. 52
  3.5.1 Osteoporosis ........................................ 53
  3.5.2 Osteoarthritis ..................................... 54
  3.5.3 Proximal Femur Fractures ......................... 55
3.6 Statistical Modelling of Femur Morphology ............... 56
  3.6.1 Segmentation ........................................ 57
  3.6.2 Model Generation ................................... 59
  3.6.3 Clinical Prediction .................................. 62
3.7 Limitations of Previous Works ........................... 63

4 General Theory ......................................... 67
  4.1 Principal Component Analysis ........................ 69
    4.1.1 Computation ...................................... 69
  4.2 Piecewise Parametric Meshes .......................... 73
    4.2.1 Quartic Lagrange Basis Function ................ 75
    4.2.2 Comparison to PDMs and Other Parametric Meshes 77
  4.3 Mesh Fitting .......................................... 79
    4.3.1 Data Error ......................................... 79
    4.3.2 Smoothing Errors .................................. 81
    4.3.3 Fitting Methods ................................... 84
  4.4 Active Shape Modelling ................................ 87
    4.4.1 Training ........................................... 87
    4.4.2 Segmentation ...................................... 88
  4.5 Cortical Thickness Mapping ............................ 93
    4.5.1 Model Fitting ..................................... 94
    4.5.2 Weighting Function ............................... 94
    4.5.3 Inner and Outer Cortical Surface Segmentation 96
  4.6 Radial Basis Functions ................................ 97
## CONTENTS

4.6.1 Common Functions ............................................. 98
4.6.2 Gaussian RBF Width ........................................... 99
4.6.3 Fitting Methods ................................................ 99
4.6.4 Evaluation Methods ............................................. 100
4.7 Summary .......................................................... 100

5 VIFM CT Images ...................................................... 101
  5.1 Background ....................................................... 102
  5.2 Ethical Approval ................................................ 102
  5.3 Image Description ................................................. 104
    5.3.1 Imaging Setup .............................................. 105
    5.3.2 Image Artefacts ............................................. 106
  5.4 Data Demographics .............................................. 108
  5.5 Use of the VIFM Dataset ......................................... 109
    5.5.1 Excluded Images ............................................. 111
    5.5.2 Training and Validation Set ................................. 111
  5.A MFC Data ........................................................ 117
  5.B X-ray CT ........................................................ 118
    5.B.1 X-ray Imaging Theory ..................................... 118
    5.B.2 Tomography .................................................. 119
    5.B.3 Projection Acquisition ..................................... 121
    5.B.4 Quantitative CT ............................................. 123

6 Femur Shape Model Training ...................................... 125
  6.1 Corresponding Regions .......................................... 126
    6.1.1 Methods ..................................................... 127
    6.1.2 Resulting Regions ......................................... 130
    6.1.3 Finalising Regions ......................................... 133
  6.2 Femur Mesh Design ............................................... 133
  6.3 Shape Model Training ........................................... 135
    6.3.1 Fitting Region Boundary Curves ........................... 136
    6.3.2 Fitting Region Meshes ..................................... 138
6.3.3 Iterative Fitting-Training .................................................. 140
6.3.4 Ensemble Mesh Training .................................................. 140
6.3.5 Training Results .......................................................... 140
6.4 Comparison to Non-regional Shape Model .......................... 143
   6.4.1 Non-Regional Shape Model .......................................... 144
   6.4.2 Comparisons .......................................................... 144
   6.4.3 Results and Discussion .............................................. 145
6.5 Conclusions and Future Work .......................................... 147
6.A Region Growing and Merging Algorithm ............................ 151
6.B Correspondent Region Sensitivity Analysis ......................... 153
6.C Addition Correspondent Region Experiments ....................... 154
   6.C.1 Cortical Thickness .................................................. 154
   6.C.2 Mean and Gaussian Curvature ..................................... 156
7 Cortex Meshing ............................................................... 159
   7.1 Introduction ............................................................. 159
   7.2 Segmentation and Meshing Process ................................ 161
      7.2.1 ASM Segmentation ............................................... 162
      7.2.2 Cortical Thickness Mapping on Mesh ......................... 166
      7.2.3 Inner and Outer Cortical Surfaces .......................... 168
   7.3 Validation Experiment ................................................ 170
      7.3.1 Leave-One-Out ................................................... 171
      7.3.2 Cortical Surface Meshes ........................................ 171
   7.4 VIFM Scans ............................................................... 178
      7.4.1 Results ............................................................. 178
      7.4.2 Enlarging ASM Training Set ................................... 182
   7.5 Discussion ............................................................... 182
      7.5.1 Training Set Size ................................................ 182
      7.5.2 Suitability of ASM ............................................... 183
      7.5.3 Consequences of the Mesh ..................................... 183
      7.5.4 Computational Cost ............................................. 184
   7.6 Conclusions ............................................................. 184
# List of Figures

1.1 Overview of the CT-to-model pipeline. ........................................... 4
1.2 Examples of cortex and cancellous morphology models. ...................... 5

2.1 Point distribution model example. .................................................. 11
2.2 Piecewise parametric mesh example. .............................................. 12
2.3 Medial descriptor example. ............................................................ 13
2.4 Statistical deformation model example. .......................................... 14
2.5 Global descriptor example. ............................................................. 14
2.6 Alignment coordinate systems for the femur. .................................... 16
2.7 PCA on 2-D data. ........................................................................... 18
2.8 Illustration of point-to-point correspondence. .................................... 23
2.9 Illustration of parametric correspondence. ........................................ 24
2.10 Illustration of population-based correspondence optimisation. ............ 26

3.1 The femur in the lower limbs. ............................................................ 32
3.2 Whole femur anatomy. .................................................................... 33
3.3 Proximal femur anatomy. ................................................................. 34
3.4 Femoral shaft anatomy. ................................................................. 35
3.5 Distal femur anatomy ....................................................................... 36
3.6 Cross section of the proximal femur. ................................................. 38
3.7 Schematic of the micro-structure of bone. ......................................... 39
3.8 Femur lengths. .................................................................................. 43
3.9 Femur widths. .................................................................................. 44
3.10 Femur neck-shaft angle. ................................................................. 45
3.11 Trabecular margin length. ............................................................... 46
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.12</td>
<td>DXA and QCT images</td>
<td>46</td>
</tr>
<tr>
<td>3.13</td>
<td>Cortical thickness versus age</td>
<td>49</td>
</tr>
<tr>
<td>3.14</td>
<td>Gender differences in age-related vBMD loss</td>
<td>50</td>
</tr>
<tr>
<td>3.15</td>
<td>CT image of osteoporotic femur</td>
<td>53</td>
</tr>
<tr>
<td>3.16</td>
<td>CT image of osteoarthritic femur</td>
<td>54</td>
</tr>
<tr>
<td>3.17</td>
<td>CT image of a cervical hip fracture</td>
<td>55</td>
</tr>
<tr>
<td>3.18</td>
<td>ASM segmentation of knee-joint bones</td>
<td>57</td>
</tr>
<tr>
<td>3.19</td>
<td>Pelvis and femur segmentation</td>
<td>58</td>
</tr>
<tr>
<td>3.20</td>
<td>Statistical model reconstruction of femur stress and strain</td>
<td>60</td>
</tr>
<tr>
<td>3.21</td>
<td>Femur reconstruction from ultrasound data</td>
<td>61</td>
</tr>
<tr>
<td>3.22</td>
<td>Femur morphologies with high risk of fracture</td>
<td>62</td>
</tr>
<tr>
<td>4.1</td>
<td>Overview of the CT-to-model pipeline</td>
<td>68</td>
</tr>
<tr>
<td>4.2</td>
<td>PCA on 2-D data</td>
<td>70</td>
</tr>
<tr>
<td>4.3</td>
<td>PCA reconstruction of 2-D data</td>
<td>71</td>
</tr>
<tr>
<td>4.4</td>
<td>A quartic-Lagrange piecewise parametric mesh</td>
<td>73</td>
</tr>
<tr>
<td>4.5</td>
<td>Quartic Lagrange line element</td>
<td>75</td>
</tr>
<tr>
<td>4.6</td>
<td>Quartic Lagrange 2-D quad element</td>
<td>76</td>
</tr>
<tr>
<td>4.7</td>
<td>Quartic Lagrange 2-D triangular element</td>
<td>76</td>
</tr>
<tr>
<td>4.8</td>
<td>Flexibility of quartic Lagrange elements</td>
<td>77</td>
</tr>
<tr>
<td>4.9</td>
<td>Data projection distance</td>
<td>79</td>
</tr>
<tr>
<td>4.10</td>
<td>Data point to material point distance</td>
<td>80</td>
</tr>
<tr>
<td>4.11</td>
<td>Material point to data point distance</td>
<td>81</td>
</tr>
<tr>
<td>4.12</td>
<td>Mesh-fitting defects</td>
<td>82</td>
</tr>
<tr>
<td>4.13</td>
<td>Inter-element smoothing</td>
<td>83</td>
</tr>
<tr>
<td>4.14</td>
<td>Free-form deformation of a rectus femoris muscle</td>
<td>85</td>
</tr>
<tr>
<td>4.15</td>
<td>Principal component of proximal femur shape</td>
<td>85</td>
</tr>
<tr>
<td>4.16</td>
<td>Illustration of image profile sampling for ASM training</td>
<td>88</td>
</tr>
<tr>
<td>4.17</td>
<td>Surface prediction along a sampled profile in ASM segmentation</td>
<td>89</td>
</tr>
<tr>
<td>4.18</td>
<td>Error correction for ASM segmentation</td>
<td>91</td>
</tr>
<tr>
<td>4.19</td>
<td>Cortical profile modelling</td>
<td>93</td>
</tr>
<tr>
<td>4.20</td>
<td>CTM profile weighting</td>
<td>95</td>
</tr>
</tbody>
</table>
List of Figures

4.21 Thin-plate spline RBF approximation of a 2-D function. . . . . . . . . . . . 98

5.1 A Typical VIFM image. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 103
5.2 The Toshiba Aquillion 16 MDCT scanner at the VIFM. . . . . . . . . . . . 104
5.3 The VIFM CT scanning room. . . . . . . . . . . . . . . . . . . . . . . . . 105
5.4 Artefacts found in VIFM images. . . . . . . . . . . . . . . . . . . . . . . . 107
5.5 Field-of-view cropping artefact. . . . . . . . . . . . . . . . . . . . . . . . . 108
5.6 Age distribution of VIFM data by gender. . . . . . . . . . . . . . . . . . . 110
5.7 Weight distribution of VIFM data, by gender. . . . . . . . . . . . . . . . . 110
5.8 Height distribution of VIFM data, by gender. . . . . . . . . . . . . . . . . 111
5.9 Examples of excluded VIFM images. . . . . . . . . . . . . . . . . . . . . . . 112
5.10 Segmentation of the training and validation set images. . . . . . . . . . . . 113
5.11 MFC proximal femur, mounted for laser scanning. . . . . . . . . . . . . . . 117
5.12 Diagram of X-ray attenuation phenomena. . . . . . . . . . . . . . . . . . . 119
5.13 Illustration of the Radon transform. . . . . . . . . . . . . . . . . . . . . . . 120
5.14 Key components of a MDCT source-detector ring. . . . . . . . . . . . . . . 122
5.15 The Mindways QCT phantom, boxed in red. . . . . . . . . . . . . . . . . . 124

6.1 Overview of the femur shape model training process. . . . . . . . . . . . . . 126
6.2 Overview of correspondent region identification. . . . . . . . . . . . . . . . . 127
6.3 Partitioning regions on the femoral surface. . . . . . . . . . . . . . . . . . . 128
6.4 Gaussian curvature sign regions on a typical femur. . . . . . . . . . . . . . 131
6.5 Correspondent regions on eight femurs. . . . . . . . . . . . . . . . . . . . . 132
6.6 Finalised correspondent regions. . . . . . . . . . . . . . . . . . . . . . . . . 133
6.7 The ensemble femur mesh. . . . . . . . . . . . . . . . . . . . . . . . . . . . 134
6.8 One-to-one correspondence of nodes across adjacent mesh boundaries. . . . 135
6.9 Overview of the ensemble and region shape model training process. . . . . . 135
6.10 Correspondent region boundary curves. . . . . . . . . . . . . . . . . . . . . 137
6.11 Region mesh fitting, illustrated using the lateral condyle. . . . . . . . . . . 139
6.12 Mesh errors and shape model compactness over fitting-training iterations. . . 141
6.13 Principal components of the ensemble femur shape model. . . . . . . . . . . 143
6.14 Cumulative principal component powers for femur shape models. . . . . . . 146
List of Figures

6.15 The region growing process. ................................................. 152
6.16 Dunn’s Index for regions cluster .......................................... 153
6.17 Cortical thickness regions on a typical femur .......................... 155
6.18 Anatomic definition of the proximal and distal femurs ............... 156
6.19 Mean and Gaussian curvature regions on a typical femur .......... 157

7.1 Cortical bone of the proximal femur ................................. 160
7.2 Deformation of the femur mesh throughout cortical surface segmentation ............................ 162
7.3 Overview of femur ASM segmentation .................................. 163
7.4 Stages of ASM segmentation ............................................. 164
7.5 Error correction for ASM segmentation ............................... 166
7.6 Cortical thickness mapping .............................................. 167
7.7 Mesh customisation using CTM data .................................... 168
7.8 Cortical mesh construction by inward projection ....................... 169
7.9 User interface of Stradwin .................................................. 170
7.10 Average cortical surface meshes ........................................ 172
7.11 Stradwin ground-truth limitations ..................................... 174
7.12 Cortical surface mesh cross sections .................................. 176
7.13 Cortical surface mesh cross sections .................................. 177
7.14 Examples of A-rated cortex meshes .................................... 179
7.15 Examples of B-rated cortex meshes .................................... 180
7.16 Examples of C-rated cortex meshes .................................... 181

8.1 Cancellous volume of the proximal femur ............................. 192
8.2 Overview of the cancellous bone registration process ............... 193
8.3 Cancellous volume registration using RBFs ............................ 195
8.4 Effects of knot spacing and width on registration .................... 196
8.5 Effect of knot spacing and width on registration error .............. 197
8.6 FFD registration of RBF-registered cancellous volume .......... 200
8.7 Registration metrics comparing RBF registration to FFD registration .................................. 202
8.8 Proximal femur composites of registered images ................. 204
8.9 Distal femur composites of registered images ........................ 205
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.10</td>
<td>Principal component variance of cancellous volume images</td>
<td>207</td>
</tr>
<tr>
<td>8.11</td>
<td>Mean cancellous volume images, and principal components of variation</td>
<td>209</td>
</tr>
<tr>
<td>9.1</td>
<td>Preliminary data from automatic morphometry measurements on the VIFM images. Femoral head diameter and femoral length distributions for males and females are as expected</td>
<td>220</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Summary of dimension reduction methods. .......................... 20

3.1 Mean femur morphometric measurements. .......................... 42

4.1 Common radial basis functions. ........................................... 98

5.1 VIFM CT scanner parameters. ........................................... 106

5.2 Average age of death, weight and height for the VIFM dataset. Standard deviation is given in brackets. .......................... 109

5.3 Mindways QCT phantom densities. ....................................... 123

6.1 Correspondent region clusters. ............................................ 131

6.2 Average mesh fitting errors. .............................................. 142

6.3 Leave-one-out approximation errors of the regional and non-regional shape models. .............................................. 147

6.4 Cortical thickness region clusters. ...................................... 154

6.5 Mean and Gaussian curvature region clusters. ......................... 156

7.1 Cortex segmentation validation errors. ................................. 173
# Glossary of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-D</td>
<td>One-Dimensional.</td>
</tr>
<tr>
<td>2-D</td>
<td>Two-Dimensional.</td>
</tr>
<tr>
<td>3-D</td>
<td>Three-Dimensional.</td>
</tr>
<tr>
<td>aBMD</td>
<td>areal Bone Mineral Density.</td>
</tr>
<tr>
<td>ASM</td>
<td>Active Shape Modelling.</td>
</tr>
<tr>
<td>BMD</td>
<td>Bone Mineral Density.</td>
</tr>
<tr>
<td>CCA</td>
<td>Canonical Correlation Analysis.</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography.</td>
</tr>
<tr>
<td>CTM</td>
<td>Cortical Thickness Mapping.</td>
</tr>
<tr>
<td>DXA</td>
<td>Dual-energy X-ray Absorptiometry.</td>
</tr>
<tr>
<td>FDA</td>
<td>Fisher’s Discriminant Analysis.</td>
</tr>
<tr>
<td>FFD</td>
<td>Free-Form Deformation.</td>
</tr>
<tr>
<td>GPA</td>
<td>General Procrustes Analysis.</td>
</tr>
<tr>
<td>HU</td>
<td>Hounsfield Unit.</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear Discriminant Analysis.</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>MDCT</td>
<td>Multi-Detector Computed Tomography.</td>
</tr>
<tr>
<td>MDL</td>
<td>Minimum Description Length.</td>
</tr>
<tr>
<td>MFC</td>
<td>Melbourne Femur Collection.</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging.</td>
</tr>
<tr>
<td>PC</td>
<td>Principal Component.</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis.</td>
</tr>
<tr>
<td>PDM</td>
<td>Point Distribution Model.</td>
</tr>
<tr>
<td>PLSR</td>
<td>Partial Least-Squares Regression.</td>
</tr>
<tr>
<td>PPM</td>
<td>Piecewise Parametric Mesh.</td>
</tr>
<tr>
<td>QCT</td>
<td>Quantitative Computed Tomography.</td>
</tr>
<tr>
<td>RBF</td>
<td>Radial Basis Function.</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Square.</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation.</td>
</tr>
<tr>
<td>SDM</td>
<td>Statistical Deformation Model.</td>
</tr>
<tr>
<td>SPHARM</td>
<td>Spherical Harmonics.</td>
</tr>
<tr>
<td>vBMD</td>
<td>volumetric Bone Mineral Density.</td>
</tr>
<tr>
<td>VIFM</td>
<td>Victorian Institute of Forensic Medicine.</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

This thesis presents an automatic pipeline for segmenting and modelling human femur morphology. Morphology encompasses femur shape, size, cortical bone thickness, and bone mineral density (BMD) distribution. The pipeline employs statistical shape analysis (SSA) methods to extract, model, and analyse these aspects from clinical X-ray Computed Tomography (CT) images. Development and validation of the pipeline was performed on a population of 262 CT images from the Victorian Institute of Forensic Medicine (VIFM).

This chapter will present the motivations for this work, address its main aims and scope, and present an overview of its key steps, which is reflected in the structure of this thesis.

1.1 Motivations

Understanding and making predictions using femur morphology has a long history in anthropology, forensics, and clinical medicine. Traditional measurement of lengths, widths and angles have been used to predict fracture risk, diagnose bone diseases, and determine age at death, amongst other applications. Until recently, however, progress has been limited by two factors: accurate and efficient methods to mathematically describe 3-D femur morphology, and sufficient sample size for statistical robustness.

The proliferation of CT imaging in the last 20 years means that there is a wealth of femur morphology data, provided that morphology can be efficiently extracted from
the images. Manual measurements lack the details of a full 3-D model, and manual segmentation is infeasible for the hundreds or more 3-D images required for statistical analysis.

Fortunately over the same period, the field of SSA has used computer graphics and machine learning methods to make sense of the vast amount of 3-D imaging data. SSA enables statistical classification and prediction using precise 3-D geometry and robust image segmentation, and thus overcomes the aforementioned limitations of accurate description and sample size. SSA has produced numerous techniques for segmentation and analysis, applied to various bones and soft-tissues. However, not all techniques are compatible, either with each other, or with femur morphology. There is thus a need for a full SSA-based pipeline, from segmentation to modelling for analysis, tailored for 3-D femur morphology.

The development of such a pipeline has been made possible by 262 CT images from the VIFM, which collects full-body CT images for its mortuary facility. This substantial set of images is larger than most used by SSA studies, so it offers a unique opportunity to develop and test CT-to-model methods. The potential of the thousands of images in the VIFM database, and other databases like it, will only be realised through the use of such an automatic segmentation and modelling system.

1.2 Central Aims and Scope

The general aim for this thesis project was to combine SSA methods for image segmentation, 3-D morphology description, and statistical analysis in a pipeline tailored for 3-D femur morphology presented in clinical CT images, in particular the VIFM images.

Specifically, the aims were:

- Automatic image segmentation
- Automatic meshing of femoral cortex, and measurement of cortical thickness
- Automatic registration of cancellous bone BMD images

The scope of the project was limited to:
1.3 NOVEL CONTRIBUTIONS

- Imaging characteristics presented by the VIFM images
- Morphological features visible in the VIFM images (e.g. individual trabeculae could not be resolved, so trabecular micro-structure was not considered)
- Normal femurs only (deformed, diseased and damaged femurs were be excluded)

The focus was on developing and assembling for the pipeline. Using the pipeline for clinical or anthropological studies was considered for future work.

1.3 Novel Contributions

As will be discussed in section 3.7, previous work in creating statistical models of femur morphology have four main limitations for which this work provides novel contributions:

1. The population size of previous works is often limited, especially for the whole femur. In the VIFM dataset used for this work, there are 262 images from which 217 full femur models are automatically created.

2. There is a lack of a complete system for image segmentation, morphology modelling, and statistical analysis, which would streamline data collection. The pipeline proposed in this thesis includes all these components.

3. There does not exist a femur description that recognises cortical and cancellous bones as separate regions. The proposed pipeline produces a 3-D cortical model with cortical bone thickness measurements, and a model of cancellous BMD distribution. The two models can be treated independently or combined for statistical analysis.

4. All previous work for the femur have used dense linear meshes, or registered images to model the femur, which results in a very large number of parameters for statistical analysis. This thesis proposes a region-based high-order piecewise parametric mesh description of femur shape that is compact and accurate.

1.4 Thesis and Pipeline Overview

As illustrated in figure 1.1, the CT-to-model pipeline is composed of two phases: training and processing. The training phase involves designing a femur mesh and training a
Figure 1.1: Main steps of the automatic femur morphology extraction pipeline, and their relevant chapters.
1.4. THESIS AND PIPELINE OVERVIEW

femur statistical shape model. The processing phase involves using that shape model to drive automatic segmentation, and modelling the various aspects of femur morphology.

Morphology modelling is based on femur architecture, which has two distinct regions: the cortex, and the internal region containing cancellous bone. As illustrated in figure 1.2, the cortex is modelled using piecewise parametric meshes, while cancellous bone is modelled as a registered quantitative-CT image. The two regions are separated because surface correspondence is easier to achieve through cortical surface landmarks. Volumetric correspondence is complicated by highly variable internal volumetric features. Therefore, the cortex is decoupled to optimise its correspondence independent of the interior volume.

To summarise, the full pipeline is as follows:

1. Input: population of femur CT images
2. Training:
   a) Design the femur mesh
b) Train a statistical shape model of the femur

3. Processing:
   a) Segment the cortex, map cortical thickness, mesh the cortex
   b) Segment and register cancellous BMD image

In the following chapters, this thesis will provide details regarding the background, implementation, and validation of the steps above. A summary of the chapters is given below.

**Background: Chapters 2 and 3**

Since the pipeline deals intimately with statistical shape analysis and femur morphology, the next two chapters will introduce these two topics, and review the current literature.

**General Theory: Chapter 4**

The general theories behind the methods used in the pipeline will be covered in this chapter. Specifically, the topics are

- Principal component analysis
- Piecewise parametric meshes
- Mesh fitting
- Active shape modelling
- Cortical thickness mapping
- Radial basis functions

This chapter is intended as a reference for readers unfamiliar with any of the topics.

**VIFM CT Images: Chapter 5**

The pipeline was developed and tested on 262 CT images collected from the VIFM. Chapter 5 provides background on the demographics of the population, the imaging characteristics, and the challenges it presented to the CT-to-model pipeline.
Training: Chapter 6

Chapter 6 covers both the design of the femur mesh, and training of the femur shape model. The chapter presents a novel method for finding correspondent regions of the femur, using surface curvature and cluster analysis. These regions were then used to design the femur mesh, which was fitted to the regions to train the shape model.

Processing: Chapters 7 and 8

The processing phase of the pipeline is described and validated in these two chapters. Chapter 7 deals with automatic segmentation and meshing of the inner and outer cortical surfaces, using a novel combination of active shape model segmentation and cortical thickness mapping. Chapter 8 describes how the inner cortical surface mesh was used to segment and register the cancellous BMD image, using a novel application of radial basis function registration.

Conclusions and Future Work: Chapter 9

Chapter 9 concludes the thesis by summarising the design and implementation of the pipeline, as well as its performance on the VIFM images. Finally, it will discuss future work regarding improving and applying the pipeline for musculoskeletal research.
Chapter 2

Statistical Shape Analysis

Statistical shape analysis (SSA) is the use of statistical methods to model variations in shape. It provides the means to quantify shape differences, generate representative shapes, or correlate shape to other measurements. Shape can be strictly defined as geometric information invariant to translation, rotation and scaling \cite{Dryden1998}. However, many SSA methods can be generalised, and are applied to data containing more general morphological information such as size and appearance.

Since the pioneering work of \cite{Thompson1917}, SSA has been widely employed to study the variations in biological specimens. The seminal work of \cite{Cootes1994} popularised the use of SSA with medical image analysis. It has provided ways to not only quantify the shape variations of objects in images, but also automatically and robustly segment the objects themselves. This chapter will review the theory and fundamental methods of SSA, while the next chapter will review the applications of SSA to femur morphology.

Key steps in SSA are:

- Describing each object in a \textit{training set} using a shape descriptor;
- Aligning the descriptions to remove translational, rotational, and size variations;
- Statistical analysis and modelling of the variations in the aligned shape descriptions.

The shape descriptor is a mathematical way to encode shape information, and is dependent on what aspects of an object’s shape are of interest. In section 2.1, a range
of descriptors for shape and morphology are reviewed.

Removal of translational, rotational, and scale variation between objects is required to obtain pure shape information. Methods for doing this are reviewed in section 2.2.

Statistical analysis commonly involves decomposing variations in shape information into modes or components. The components can then be used for shape generation and classification. Details of these statistical methods are reviewed in section 2.3.

One of the most important considerations in SSA is ensuring correspondence in the shape description. Each parameter in the shape descriptor should describe the same landmark for all shape instances, so that the only variation in a parameter is due to shape. Some methods for achieving correspondence iterates through the three main steps above. Correspondence methods will thus be reviewed last, in section 2.4.

2.1 Shape Descriptions

The shape description is responsible for mathematically describing the shape, and in some cases, other spatially varying quantities, of an object. Shape descriptions range in complexity from simple hand-placed landmark points, to parametric meshes, to wavelet coefficients. Whatever the description, its parameters must be suitable for statistical analysis. The choice of description depends not only on its flexibility in capturing morphology, but also the intended statistical analysis, and methods available for ensuring correspondent descriptions. The first two factors are discussed for each of the commonly used descriptors presented below. Correspondence methods are often not specific to a particular descriptor, and so will be covered later in section 2.4.

2.1.1 Point Distribution Models (PDMs)

One of the simplest descriptions of an object’s shape is a distribution of points at definable landmarks (Fig 2.1). The coordinates of each landmark form the variables for statistical analysis. When landmarks are manually placed, there is a practical limit on the number of landmarks feasible (O’Higgins 2000; O’Higgins and Dryden 1993). However, as the number of landmarks increase, so does the detail to which shape is described - a necessary trade-off. Shape descriptions based on a dense distribution of
2.1. SHAPE DESCRIPTIONS

11

Figure 2.1: A 2-D point distribution model. The shape is described by the coordinates of points (landmarks) around the object boundary.

points are termed Point Distribution Models (PDM) in Cootes et al. (1992, 1994). The PDM is the most popular shape description (Heimann and Meinzer, 2009). This is partly due to the ease with which dense point-cloud surfaces can be generated, and partly because correspondence methods for PDMs have developed in parallel (see Sec. 2.4). The number of points in a PDM of a 3-D object is typically in the thousands, or tens of thousands. It is easy to generate these points on one particular shape by simply discretising the surface. However, corresponding the points between shapes is a challenging problem with continuing research.

2.1.2 Piecewise Parametric Meshes (PPMs)

Piecewise parametric functions can be used to represent continuous curves, surfaces, and volumes, as a mesh of smooth patches or elements (Fig. 2.2). Mesh geometry is controlled by parameters, such as the coordinates, of nodes or control points within each element. In addition to geometry, the meshes can also interpolate fields related to morphology, such as thickness over a surface (Zhang et al., 2012), or bone density in a volume (Shim et al., 2007). For SSA, mesh parameters, commonly associated with mesh nodes, are used in statistical analysis. The number of parameters is typically much smaller than for PDMs, which makes PPMs more efficient, especially for smooth objects.
12

CHAPTER 2. STATISTICAL SHAPE ANALYSIS

Figure 2.2: A 2-D piece-wise parametric mesh shape description. The shape is described by piecewise parametric functions (outlined in red) controlled by parameters at nodes (black points). The mesh can also interpolate other properties in the object’s interior.

PPMs as descriptors for SSA began with curves for 2-D shapes such as gait patterns (Niyogi and Adelson, 1994; Baumberg and Hogg, 1994) and cross-sections of the intercondylar notch (Shepstone et al., 2001). For surfaces, B-spline patches were used to model the proximal tibia in Hafner et al. (2000), and the heart in Horkaew and Yang (2003) and Garcia-Barnes et al. (2010). Non-Uniform Rational Splines (NURBS) were used to model and segment kidneys in Tsagaan et al. (2002), where principal curvatures at correspondent points on the mesh were used as the shape parameters. Principal curvatures are invariant to translation, rotation, and scale, so no alignment of the surfaces were needed. In terms of meshes with more than just shape information, cortical thickness was embedded in Lagrange-element meshes of the femur in Zhang et al. (2012), and bone mineral density was embedded in linear tetrahedral meshes in Bryan et al. (2009).

2.1.3 Medial Descriptors

Medial descriptions of shape are characterised by a centreline and varying radii along its length (Fig. 2.3). The centreline can be discretised into individual medial atoms (Fritsch et al., 1997). Each atom is defined by its location, vectors describing its local frame (orientation), and vectors pointing to boundaries. These values are used in subsequent statistical analysis. Unlike simple point coordinates, as used in PDMs and
some parametric surface descriptions, these parameters do not vary linearly. Because of this, common statistical analysis applicable to PDMs cannot be used with medial descriptors.

Skeletal representation of biological shapes was first introduced by Blum (1973). Medial description, referred to as fixed-topology skeleton, was used to model the shape of the corpus callosum in Golland et al. (1999). The concept of medial atoms was introduced with deformable shape loci in Fritsch et al. (1997), for describing 2-D shapes. Medial description was extended to 3-D in Joshi et al. (2002) and Pizer et al. (2003), along with a hierarchical implementation, and termed m-reps. For an m-reps representation of a 3-D slab-like volume, medial atoms are arranged in a 2-D manifold interpolated by Bezier splines. Yushkevich et al. (2006) furthered generalised m-reps to a continuous representation, and incorporated image appearance into the description.

2.1.4 Statistical Deformation Models

Meshing an object’s surface or volume for morphological analysis is not always necessary. Many authors have simply used registered images as a descriptor of appearance, and the deformation to register the image as a descriptor of shape (Fig. 2.4). This approach is referred to as statistical deformation models (SDMs) (Rueckert et al., 2003).
Figure 2.4: A 2-D statistical deformation model. The shape is described by the deformation of a reference shape to the data shape. In this example, the deformation is represented by the displacement of the control points (black discs) of the spline grid in which the shape is embedded. Registering images means that object appearance can also be sampled.

Querol et al., 2006; Loeckx et al., 2003; Rajamani et al., 2007). Registration is often performed using 3-D B-spline free-form deformation (Schnabel et al., 2001), in which case the change in position of each B-spline control point is used for statistical shape analysis, while pixel values in the registered images can be used to analyse appearance, or density, depending on the image modality.

2.1.5 Global Descriptors

Figure 2.5: An illustration of the modes of oscillation that form a 2-D global shape description. The shape is described by the coefficients of an oscillatory function used to approximate the shape.

Global descriptors parameterise an object’s shape based on its spatial frequency
2.1. SHAPE DESCRIPTIONS

components. 2-D and 3-D shapes can be decomposed into the oscillatory modes of various basis functions of corresponding dimensions (Fig. 2.5). The coefficients for these modes are used in statistical analysis. Examples of global descriptors include Fourier decompositions (Persoon and Fu, 1977; Kuhl and Giardina, 1982), spherical harmonics (SPHARM) (Székely et al., 1995; Kelemen et al., 1999; Brechbhler et al., 1995), and Zernike moments (Canterakis, 1999; Venkatraman et al., 2009; Chen and Sun, 2010; Millan et al., 2007).

Original implementations were limited to surfaces with spherical topology, on which points must be mapped to spherical coordinates in a correspondent manner. Surface harmonics (Matheny and Goldgof, 1995) extend SPHARMs to non-spherical topologies, by mapping to non-spherical coordinate systems. The more general wavelet-based methods (Davatzikos et al., 2003; Nain et al., 2007) accommodate surfaces of arbitrary topology by partitioning the surface into patches.

The global shape descriptors described above give good representation of smooth shapes using relatively few terms, making them efficient for soft-tissue structures (Millan et al., 2007; Mangin et al., 2004). Conversely, a large number of terms are required to accurately approximate high-frequency features, such as sharp ridges. A disadvantage of the frequency-based descriptions is that their parameters do not describe local geometry, so it is less intuitive to interpret local shape changes.

2.1.6 Summary

This section has covered the main categories of descriptors for the statistical analysis of shape and morphology. PDM-based descriptors have stood the test of time, owing to their simplicity and intuitive extension from traditional landmarks. Global shape descriptors provide more compact descriptions of shape, especially on smooth surfaces, and have been popular for modelling brain structure. Medial descriptors have seen an increase in popularity in recent years with the arrival of suitable statistical methods for dealing with its non-linear parameters. Piecewise parametric surfaces combine compactness, with superior flexibility in modelling shapes of arbitrary topology. Ensuring correspondence is a problem faced by all descriptors, and solutions will be reviewed in section 2.4.
2.2 Alignment

Shape variations are invariant to translation, rotation and scale, otherwise known as similarity transforms. Given a training set of shapes, variations due to similarity transforms can be removed by aligning the shapes to a common coordinate system.

Many anatomical structures have a well-defined anatomic coordinate system based on anatomic landmarks (e.g. bones [Wu et al., 2002]). So shapes can be aligned by their anatomic coordinate systems (Fig. 2.6a). In lieu of reliable anatomic landmarks, shapes can also be aligned to their centre of mass and principal axes of inertia (Fig. 2.6b), as was done in [Dalal et al., 2007].

A more general alignment method is the Procrustes analysis. A full Procrustes analysis removes all similarity-transform variations, while a partial Procrustes analysis leaves size variations. Both versions are included in a general framework called the General Procrustes Analysis (GPA) [Gower, 1975].
2.3. **STATISTICAL ANALYSIS**

A Procrustes analysis is performed by transforming a group of geometries to align with a target geometry. When the target is known, the similarity transforms can be calculated analytically, by minimising the mean squared distance between corresponding landmarks (Goodall, 1991). Alternatively, Gower (1975) advocates iteratively refining the target geometry: following an initial Procrustes analysis, the mean shape is used as the target for the next iteration, where upon an updated mean shape is produced. This is repeated until the mean shape converges. In the absence of corresponding landmarks, iterative closest point (ICP) (Besl and McKay, 1992; Kapoutsis et al., 1999) can be used to calculated the aligning transformations.

### 2.3 Statistical Analysis

Statistical analysis is performed on the parameters of a training set of aligned shape descriptions, to find underlying components of variation. The number of components is expected to be far fewer than the number of shape description parameters due to covariance between parameters. Consequently, dimension reduction (Sec 2.3.1) is usually the first analysis performed. The shapes can then be described in terms of the reduced components, allowing correlation (Sec. 2.3.2) and classification (Sec 2.3.3) analysis to be performed efficiently.

#### 2.3.1 Dimension Reduction

Dimension reduction is necessary because the number of parameters in shape or morphology descriptions are generally too high for efficient or meaningful statistical analysis. Given a training set of shapes, dimension reduction decomposes shape variations into a few important components. The space spanned by these components is commonly referred to as the shape space. A point in the shape space is a particular combination of the components, and therefore an instance of shape.

Principal component analysis (PCA) is the most common dimension reduction method for shape analysis (Heimann and Meinzer, 2009). PCA decomposes a training set of observations (e.g. shape descriptions) into a mean, and orthogonal principal components, ranked in descending variance. Given $n$ shapes, each described by $m$
parameters, they can be assembled into a $m \times n$ data matrix. PCA is done through eigen-decomposition of the covariance matrix of the data matrix, or singular value decomposition of the data matrix itself (see section 4.1 in the next chapter for more details). Intuitively, the principal components are found by rotating orthogonal axes so that the first axis aligns with the direction of maximum variance, the second axes with the direction of second greatest variance, and so on (Fig. 2.7). Assuming meaningful variations in the training set, the first few modes will capture most of the variation. This means that shapes can be approximately reconstructed by linear combinations of just the first few modes, leading to a reduction in the dimensionality of the data.

PCA was first used to model shape variation by Cootes et al. (1992) and Kent (1994). Since then, it has become the standard shape-modelling statistical method. The popularity of PCA can be attributed to its intuitive theory, and the fact that its assumption of data lying in linear space is met by popular PDM-based shape descrip-

Figure 2.7: Illustration of PCA on 2-D data. PCA find the orthogonal axes along which variance is maximised (red lines). The data can be approximated by their projections on just the first principal component (longer red line), which reduces the dimensionality of the data.
2.3. STATISTICAL ANALYSIS

As a least-squares method, PCA is sensitive to outliers. Robust PCA (Hubert et al., 2002; Hubert and Engelen, 2004) or weighted PCA (Skocaj et al., 2007) overcomes this issue by allowing input variables to be weighted. Unreliable data such as outliers can be down-weighted to remove their effect from the analysis. In Schmid et al. (2011), the power of weighted PCA was demonstrated on shape and appearance models created from incomplete images of the pelvis. Principal components describe shape variations over the whole shape, whereas local variations are sometimes of interest. Orthomax rotation (Stegmann et al., 2006), and sparse PCA (Sjöstrand et al., 2006; Zou et al., 2006), can produce orthogonal components that describe local shape variations.

PCA assumes normality in data distribution, and that significant components of variation are orthogonal (Shlens, 2005). Both of these assumptions can be violated in a real population of shapes. Independent component analysis (ICA) (Hyvärinen et al., 2002) does not assume either, only that components are statistically independent. This makes ICA arguably more sensitive than PCA in detecting significant components of variation, and is also capable of modelling local variations. However, ICA does not rank its components in terms of significance, and the number of components must be defined a-priori. Statistical independence is defined differently by different ICA methods. The popular FastICA method (Hyvärinen et al., 2002) maximises the non-Gaussian-ness of signals, while ICA through infomax (Bell and Sejnowski, 1995) minimises mutual information between signals. Examples of the use of ICA in SSA include myocardial disease classification (Üzümçü et al., 2003; Suinesiaputra et al., 2009), and body shape analysis (Ruto et al., 2006). In comparing ICA to PCA, both Ruto et al. (2006) and Suinesiaputra et al. (2009) concluded that independent components expressed more local variations than principal components.

PCA and ICA assumes that its data lies in linear, or Euclidean space. However, this assumption is violated when nonlinear variations exist in the population, or when parameters of the shape or morphology description do not lie in Euclidean space. The former case is common for biological structures, while the latter is true for some shape descriptors (e.g. m-reps). Kernel PCA (Twining and Taylor, 2001) copes with nonlinear data, and has been used for shape analysis (Samuel et al., 2006), and shape
model-drive image segmentation (Yogesh et al., 2006). In an analysis of spine shape by Boisvert et al. (2008), non-linear descriptors of vertebral shape and orientation were modelled using the Fréchet mean (Le, 2001) and a generalised covariance matrix. In Pizer et al. (2003), shape variations encoded by m-reps were modelled by a Markov model. Alternatively, principal geodesic analysis was proposed for m-reps in Fletcher et al. (2004) and Dam et al. (2008). A summary of the dimension reduction methods above can be found in table 2.1.

### 2.3.2 Correlation

Shape correlations between objects allow shape prediction or generation on a group of objects, such as bones in a joint. Yang et al. (2008) used canonical correlation analysis (CCA) (Hardoon et al., 2004) and partial least-squares regression (PLSR) (Wold, 1975) to model the shape correlations between the humerus and the scapula. Given two sets of parameters, CCA seeks a set of components (or basis) along which the two sets of parameters are maximally correlated. PLSR performs a similar function, but aims to maximise covariance between parameter sets instead (Sun et al., 2009; Rosipal and Krämer, 2006).

Correlations can also be made between shape and surrogate measurements, such as bone length, or body stature. Blanc et al. (2012) performed these correlations for

---

**Table 2.1: Summary of dimension reduction methods used in literature for SSA**

<table>
<thead>
<tr>
<th>Method</th>
<th>Assumes Linearity</th>
<th>Orthogonal Components</th>
<th>Ranks Components</th>
<th>Global/Local Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>global</td>
</tr>
<tr>
<td>Robust PCA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>global</td>
</tr>
<tr>
<td>Orthomax Rotation</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>local</td>
</tr>
<tr>
<td>Sparse PCA</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>local</td>
</tr>
<tr>
<td>Kernel PCA</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>global</td>
</tr>
<tr>
<td>ICA</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>local</td>
</tr>
</tbody>
</table>


2.3. STATISTICAL ANALYSIS

the femur and tibia, using a number of different regression methods. Between CCA and PLSR, no clear difference in accuracy was found. The authors advocated the use of simple, linear methods such as PLSR, rather than more complex methods such as CCA or nonlinear kernel models. As discussed in Wegelin et al. (2000), PLSR offers two significant advantages over CCA: numerical stability; and ranking of components in terms of significance. The first factor makes PLSR more suitable for shape analysis, given that there are typically far more shape description parameters than training shapes. The second factor means that the relations produced by PLSR are clearer and easier to interpret.

2.3.3 Classification

It is often useful to compare differences in shape between two populations, for example, male–female, or normal–pathological. This is done by performing a discriminant analysis which, in addition to finding the difference between two populations, also trains a classifier that predicts the classification of a new shape. To avoid high-dimensional data, these classification analyses are carried out on dimension-reduced shape data.

Linear discriminant analysis (LDA) (Klecka, 1980), and Fisher’s discriminant analysis (FDA) (Fisher, 1936), are tests for significant differences between two groups. FDA differs from LDA in not assuming normally distributed data or equal covariance between populations. Typically, the two groups of data undergo dimension reduction by PCA on the combined set of data. Then, LDA or FDA is performed on the projections of the combined set on a reduced number of principal components. Both methods try to find a hyperplane in the reduced space that best separates the two groups of data, such that the difference between the two groups is described by the vector normal to the plane. Shepstone et al. (2001) used FDA to find the principal mode that contained the greatest inter-group variation for normal and arthritic femur intercondylar notches. Mahfouz et al. (2009) used FDA to calculate a mode describing male–female variation in femoral shape. FDA was also used in Golland et al. (1999) to differentiate between normal and abnormal corpus callosum structure.
2.4 Shape Description Correspondence

The description of set of shapes or morphologies must be correspondent for meaningful statistical analysis. Due to the large number of parameters needed to describe shape and morphology in 3D, manually assigning correspondence is impractical, and inaccurate \cite{Styner2003}. Thus many automated methods have been developed to ensure correspondence in descriptors. Correspondence is one of the major factors influencing model quality, but is also the most challenging aspect.

Correspondence methods can generally be divided into four broad categories. The first category is for methods that try to find point-to-point correspondence between two point clouds. Due to the popularity of PDMs, this category is the most widely used. The second category covers more general methods that matches the parameterisation, or mapping, of two surfaces to each other, or a reference, and can be used for descriptors beyond PDMs. Methods of the two categories above tend to establish correspondence between two objects. In the third category, correspondence is optimised, taking into account the whole training set. Given the plethora of shape descriptions and correspondence methods, information theoretic approaches can provide quantitative comparisons for selecting the optimal model.

2.4.1 Point-to-Point

Point to point methods try identify corresponding points (landmarks) between two dense, unstructured clouds of surface points (or vertices or a mesh). This is a common method of establishing correspondence for PDMs. The two surfaces are not assumed to have the same number of points. One surface can be designated the reference, and the other the data. For every point on the reference surface, the method finds a corresponding point on the data surface.

The most straightforward method of doing this is to register the two point clouds, and for each reference point, assign its closest data point as its correspondent point \cite{Hunter1992, Subsol1998, Kapoutsis1999, Barratt2008}, as illustrated in figure 2.8. The correspondent points in the data point cloud then become the landmarks of the data surface. Surface curvature, normal direction and local image
2.4. SHAPE DESCRIPTION CORRESPONDENCE

Figure 2.8: A simple example of finding point-to-point correspondence for a point distribution model. The reference (black) and data boundaries (blue) are aligned, then the closest data point to each reference point is found as its correspondent point, as indicated by the red arrows. Local curvature and other information can be used to modify the distance measure, and improve the robustness.

appearance can also augment the distance measure (Caunce and Taylor, 2001; Shen et al., 2001; Wang et al., 2005), so that correspondence is not solely based on distance. The matching of points is further explored in Hufnagel et al. (2008), where probabilistic one-to-many correspondence is established between target and source points.

The methods above assign correspondence between existing points. An alternative is to create new points, or modify existing points, to create correspondence. Wang et al. (2000) presents a method for distributing correspondent points between a small number of seed landmarks, using geodesic distances from the seed landmarks, and surface curvature as guides. In Novotni et al. (2005), features, or “salient points”, are created on both reference and data surfaces based on curvature. Then, a correspondent subset of these salient points are picked as landmarks, which are the set of points that minimises the bending energy in their warping from reference to data. Dalal et al. (2007) slides source points along the surface to minimise two objectives: the sum of squared distances between data and reference points, and the thin-plate bending energy.
required to deform to data to the reference surface.

2.4.2 Parameterisation

Domain parameterisation considers the problem of mapping boundaries, surfaces, or volumes to a reference space. An example is mapping surfaces in 3-D space to 2-D parametric space as for parametric surfaces (see Sec. 2.1.2). If two surfaces have correspondent parameterisation, then a point defined in their common parametric space (sometimes referred to as a material point) will map to correspondent points on each surface.

![Figure 2.9: A simple example of parameterisation correspondence. Both data and reference piecewise-parametric meshes have the same shape, but their interiors are differently mapped, or parameterised, by the elements. The goal is to modify the data parameterisation to match that of the reference.](image)

There are two approaches to obtaining correspondent parameterisations between the reference and data. The first is to smoothly deform the reference to fit to the data, so that the deformed reference parameterises the data. The second is to adjust the parameterisation of the data surface to match that of the reference (Fig. 2.9).

The first approach is used in Allen et al. (2003) and Bryan et al. (2009) to fit reference triangle meshes of the human body and femur, respectively, to data surfaces. A coarse to fine fitting procedure, and mesh smoothness constraints ensure a smooth deformation of the reference mesh. In Querol et al. (2006), Whitmarsh et al. (2011)
2.4. SHAPE DESCRIPTION CORRESPONDENCE

and Frangi et al. (2002), volumes are parameterised by registering a reference image to data images. This is the same approach as taken by statistical deformation models.

The second approach tends to be limited to surfaces, where reference and data surfaces are mapped to a simpler surface (or surfaces) of the same topology. Mapping to a sphere is a common choice, especially for SPHARM descriptions, so that surfaces can be parameterised by spherical coordinates (Székely et al., 1995; Brechbehler et al., 1995; Kelemen et al., 1999; Praun et al., 2001; Floater and Hormann, 2005). Of course, problems arise when surfaces are not topologically spherical. Surfaces of arbitrary topology can be parameterised by partitioning the surface into patches, which are parameterised individually to a unit disc (Zöckler et al., 2000; Lamecker et al., 2002), for example. The assumption is that if surfaces are partitioned into correspondent patches, then the parameterisation of each patch will also be correspondent across surfaces, due to each patch’s similar geometry. More sophisticated methods attempt to optimise correspondence in patches. For example, in Horkiewicz and Yang (2003), each patch is parameterised by a B-spline surface. The B-spline knots (which control parameterisation) are initially placed using harmonic mapping, then adjusted using the minimum description length (MDL) method (see Sec. 2.4.3). Another example is the work by Wang et al. (2005), in which Gaussian curvature is used to automatically partition surfaces into correspondent patches. The parameterisation within each patch is then made correspondent by maximising the mutual information of surface curvature.

2.4.3 Population-based

Correspondence can be optimised by considering the entire training set, rather than aiming for correspondence to a particular reference, which can heavily influence the result. A requirement for correspondence optimisation across a training set is some metric that measures the quality of training set-wide correspondence. In the first work of this sort by Kotcheff and Taylor (1998), the metric was the determinant of the resulting shape model’s covariance matrix. Minimising the determinant promoted the compactness of the shape model, with the first few principal components having higher eigenvalues. In other words, a model that explained more variation using fewer components. Each iteration of the optimisation involved parameterising the shapes (mapping
landmark to a unit circle), training the statistical model to calculate the covariance matrix determinant, and updating the parameterisation to minimise the determinant (Fig. 2.10). This iterative framework has become the basis of most population-based correspondence optimisations. Another important result is the use of a measure of shape model quality (compactness) as a reflection of correspondence quality.

Figure 2.10: Main steps in the population-based correspondence optimisation process, based on Kotcheff and Taylor (1998). In each iteration, a shape model is trained from the population, a metric related to shape model quality is calculated, and this metric is optimised by adjusting the parameterisation of each training shape.

Davies et al. (2002a) added generality and specificity to compactness in their measure of model quality. Generality is defined as the model’s ability to represent shape instances outside of the training set. Specificity is defined as the model’s stability in only producing valid shapes. All three measures are incorporated into an objective function based on the minimum description length (MDL) principal. MDL postulates that the best description of a set of data leads to the best compression of data. Applied to shape models, this means that the best parameterisation produces a model which describes each shape using the minimum number of principal components and additional error terms. This method was applied to landmarks on 2-D boundaries, using a genetic algorithm for the optimisation. In Davies et al. (2002b), the method was extended to
landmarks on topologically spherical 3-D surfaces, by parameterising landmarks using spherical coordinates. Thodberg and Olafsdottir (2003) added curvature matching, along with a simplified MDL algorithm, which allowed the gradient of the objective function to be derived in Ericsson and Aström (2003) for 2D, and in Heimann et al. (2005) for 3D. This allowed gradient-descent optimisation to be used, for a significant gain in computational performance. An extension of MDL to arbitrary topologies is found in Horkaew and Yang (2003), where MDL is performed independently on a number of surface patches.

MDL has become a benchmark method for optimising correspondence across a training set (Styner et al., 2003; Munsell et al., 2008). However, a number of alternative population-based methods also exist. In Hill et al. (2000) the mean shape is generated from a series of pairwise correspondence operations. These are initially performed between training set shapes, then between the means of these pairs, and so on up a binary tree, until landmarks are finally generated on the overall mean shape. These landmarks are then propagated down the binary tree to each shape in the training set. Another approach is presented in Seshamani et al. (2011), where point-wise correspondence for all training shapes is iteratively refined. In each iteration, a shape model is produced, and used to reconstruct each training shape in a leave-one-out fashion. Landmarks on each training shape are then moved in the direction their corresponding point on the reconstruction. In this way, shape model generality is optimised.

2.4.4 Model Selection

There has been limited work in quantitatively comparing the numerous shape descriptions and correspondence methods. In Styner et al. (2003), MDL was found to be superior to SPHARM and the subdivision surface method of Wang et al. (2000) in terms of generality, specificity, and compactness. In Styner et al. (2007), it was shown that the choice of shape description and correspondence method significantly affected the ability of resulting shape models to identify shape differences in brain structures. However, these were a comparison of a handful of methods amongst many. The criterion of generality, specificity, and compactness used for MDL are applicable for such

\footnote{An open source implementation can be found in Heimann et al. (2006)}
comparisons, but they ignore another factor: the number of shape description parameters. An overly complex description can introduce more noise into the resulting shape model, adversely affecting model compactness and predictive abilities. Such a behaviour is known as overfitting. To the best of the author’s knowledge, the problem of shape description overfitting has received little attention.

Information criterion methods, such as Akaike information criterion (AIC) ([Sakamoto et al., 1986], [Burnham and Anderson, 2002]), Bayesian information criterion (BIC) ([Schwarz, 1978]), and MDL, are used to measure the goodness of fit of a model, and penalise overfitting. They are widely used in the fields of machine-learning and statistics, but have seen little uptake in SSA. In the MDL method of [Davies et al., 2002a] (and its derivatives), the number of shape description parameters is held constant. But in theory, it should be possible to incorporate the number of parameters into the description length. However, this implies that decisions regarding adding or removing parameters (e.g. landmarks) be made per iteration of the optimisation. For typically thousands of parameters, the computational cost may render this idea impractical. It is still possible that information criterion can be applied to compare different shape descriptions, correspondence methods, and their resulting shape models. Beyond the scope of this thesis, the application of information criterion methods in SSA warrants further study.

2.4.5 Summary

Much work in statistical shape analysis has been to do with ensuring correspondence in shape descriptions. While most correspondence methods deal with PDM descriptions, the development of more sophisticated shape descriptors has driven correspondence methods forward. For example, the adoption of SPHARMs resulted in much work in creating correspondent parameterisations on a spherical topology. These methods, in turn, provided the basis for the MDL method. Patch-based surface descriptions allowed arbitrary topologies to be handled, and have led to patch-based correspondence optimisation.

As demonstrated by the success of MDL, it is good to infer correspondent landmarks using all training data. Older methods try to create one-to-one correspondence
between the reference and each training shape individually. Newer methods optimise correspondence across the whole training set. Some do this by optimising the resulting shape model quality, while others use the training set to identify important features to guide correspondence optimisation. Quantitative comparisons of correspondence methods (and shape descriptions) demands further investigation, especially in terms of model complexity by adoption of information criterion methods.

2.5 Final Remarks

By describing complex 3-D geometries, modern statistical shape models capture much more information than traditional morphometric measurements. This has increased the sensitivity for describing variations, and finding correlations between morphology and pathologies or anthropological measurements. However, methods for ensuring description correspondence are paramount for ensuring shape model accuracy.
Chapter 3

Femur Morphology

Femur morphology is the study of femur form, encompassing its surface shape, and internal bone structure. The extraction and modelling of these two aspects are the focus of this thesis. This chapter will present the current understanding of femur morphology and its natural and pathological variation.

The chapter begins with a review of femur anatomy in section 3.1 and bone biology in section 3.2. The ways in which femur morphology is traditionally measured is presented in section 3.3, which provides the basis for a review of natural variations in section 3.4 and abnormalities in section 3.5. The role statistical shape analysis has played in quantifying femur morphology is reviewed in section 3.6. Finally, the limitations of these prior works, and therefore the rational for this thesis, are discussed in section 3.7.

3.1 Femur Anatomy

The femur is the longest and strongest bone in the human body [White and Folkens 2005]. Its shape is generally divided into the upper extremities (proximal femur), the femoral shaft (diaphysis), and the lower extremities (distal femur). The proximal femur forms part of the hip joint, and the distal femur forms part of the knee joint. Between these two joints, the femur supports the weight of the body at rest, and in motion, which it facilitates through numerous attached muscles. Figure 3.1 shows the femur relative to other bones of the lower limbs. Dense cortical bone forms the outer layer
of the femur. Internal to the cortex, plate and rod-like trabecular bone forms the cancellous bone (spongy) structure found in the upper and lower extremities. Marrow fills the femoral shaft, and voids in the cancellous bone. The articular surfaces of the femur, at the joints, are covered by cartilage.

### 3.1.1 External Features

Besides the general shape of the femur, the most obvious features on the femoral surface are the articular joint surfaces, and muscle and ligament insertion sites. The detail to which they have been mapped is a testament to their importance, especially with respect to femur function. This section will detail these features in relation to the proximal femur, femoral shaft, and distal femur. The locations of these features are illustrated in figure 3.2.
3.1. FEMUR ANATOMY

Figure 3.2: Major external anatomic features of the femur. Coloured patches denote muscle insertions (blue) and origins (red). Originally produced by the author for this thesis.

Proximal Femur

Anatomic features and attachment sites of the proximal femur are illustrated in figure 3.3. The main anatomic features are the femoral head (epiphysis), and the proximal metaphysis, which consists of the femoral neck, the greater trochanter, and the lesser trochanter. The complex arrangement of a large number of muscles attached to the proximal femur enable the wide range of motion at the hip joint.

The femoral head meets the acetabulum of the pelvis, forming the acetabulo-femoral joint (hip joint). The femoral head is shaped as the greater part of a sphere, and is covered by smooth cartilage. The fovea capitis is a small depression on the medial aspect of the femoral head. The ligament of the femoral head connects the fovea capitis to the acetabulum.

The femoral neck extends from the femoral head, increasing in diameter, to the intertrochanteric line (anterior) and the intertrochanteric crest (posterior). Lateral
to the femoral neck is the greater trochanter, while at its inferior end is the lesser trochanter. The articular capsule of the hip joint is anchored around the base of the femoral neck.

The prominence of the greater trochanter reflects its role as the attachment point for several muscles. The tendon of the gluteus medius attaches to the lateral face, the gluteus minimus attaches to the anterior face, the piriforms to the superior ridge, and the vastus lateralis to the inferior border. To the medial surface are attached the obturator internus, superior and inferior gemellus, and the obturator externus at the trochanteric fossa.

The lesser trochanter is a conical prominence at the interface with the femoral shaft. To its apex are attached the psoas major and the iliacus, which are responsible for the flexion and external rotation of the hip joint. On the posterior side of the femur, the intertrochanteric crest runs from the lesser trochanter to the superior, posterior prominence of the greater trochanter. Halfway along the crest, the quadratus femoris attaches the quadrate tubercle. The intertrochanteric line runs between the trochanters on the anterior side of the femur.
3.1. FEMUR ANATOMY

Figure 3.4: Anatomic features (bold) and muscles (italic) of the femoral shaft. Coloured patches denote muscle insertions (blue) and origins (red). Originally produced by the author for this thesis.

Femoral Shaft

Anatomic features and attachment sites of the femoral shaft are illustrated in figure 3.4. The femoral shaft (diaphysis) begins distal to the lesser trochanter. The shaft is approximately cylindrical, while slightly broader on the anterior side. It narrows slightly from the proximal end, and broadens towards the condyles are the distal end. The shaft is slightly bowed such that it is convex on the anterior side and concave on the posterior side.

The surface of the shaft is relatively featureless, except for the linea aspera - a rough, prominent ridge running down the posterior side of the shaft. It begins in the proximal femur as two main ridges - the lateral ridge begins from the lateral base of the greater trochanter, and the smaller pectineal line runs distally from the inferior side of the lesser trochanter. The lateral ridge and pectineal line merge together about a third of the way down the femoral shaft into a single ridge with medial and lateral
lips. Inferior to the midpoint of the femur, the linea aspera diverges again into lateral and medial ridges. Each continue to the epicondyles.

The proximal lateral ridge is attached to by the gluteus maximus. Due to the size and power of this muscle, the attachment is especially prominent, and is referred to as the gluteal tuberosity. The pectineal line is attached to by the pectineus. The medial and lateral lips of the linea aspera are attached to by the vastus medialis, and lateralis, respectively. To the distal lateral and medials are attached the biceps femoris and adductor magnus, respectively. Additional muscles attached to the linea aspera are the adductor brevis, adductor longus, and the iliacus.

The popliteal surface is the triangular area bounded by the distal lateral and medial ridges, and the top of the condyles. The femoral nutrient canal is found towards the superior corner of this surface.

Distal Femur

![Figure 3.5: Anatomic features (bold) and muscles (italic) of the distal femur. Coloured patches denote muscle insertions (blue) and origins (red). Originally produced by the author for this thesis.](image)

Anatomic features and attachment sites of the distal femur are illustrated in figure 3.5. The distal femur encompasses the lateral and medial condyles (distal epiphysis), and the distal metaphysis, which consists of the epicondyles, and the adductor tubercle. The distal femur is a part of the knee joint, with cartilage covering the an-
3.1. **FEMUR ANATOMY**

terior, inferior, and posterior sides of the condyles, which form a continuous, smooth surface. The lateral and medial condyles protruded posteriorly, forming the inter-condylar fossa in between. Condylar prominence anteriorly is much less pronounced, and forms the shallow patella groove, in which the patella slides. The lateral condyle is broader, while the medial condyle projects further along the direction of the femoral shaft. However, in its natural anatomic position, the base of the two condyles are level at the knee joint.

The intercondylar fossa is not covered by cartilage, and provides attachment points for the anterior and posterior cruciate ligaments. The epicondyles are situated on the sides of the condyles, where the lateral epicondyle is the attachment site of the popliteus muscle and the fibular collateral ligament, and the medial epicondyle serves as the attachment site of the tibial collateral ligament. In addition, in the upper part of the medial epicondyle is the adducter tubercle, to which is attached the gastrocnemius.

### 3.1.2 Cortical Bone

Dense cortical bone forms the outer shell of the femur. Cortical thickness ranges from about 7 mm in the femoral shaft (Bertelsen et al., 1995), to under 0.1 mm in the femoral head (Grynpas et al., 1991). However, thickness varies greatly with age and gender (Sec. 3.4). Its stiffness and spatial distribution mean that cortical bone provides most of the femur’s resistance to bending.

The cortex is thickest in the linea aspera, and the medial and lateral sides of the femoral shaft. Thickness at the linea aspera can be explained by the attachment of major thigh muscles. Medial and lateral thickness in the shaft can be explained by the bending moment in the coronal plane, created by the cantilevered femoral head, under a standing posture load. Thicker cortex in the medial and lateral aspects compensate for the increased compressive and tensile stresses, respectively. Medial-side thickening of the cortex continues up the lower side of the femoral neck, where high compressive stresses occur.
3.1.3 Cancellous Bone

Cancellous, or spongy bone is found in the proximal and distal femur. It is a lattice of trabecular bone elements in the shape of plates, rods and struts. The lattice structure is often anisotropic. That is, there is a preferred direction in the arrangement of trabecular plates. This can be clearly seen for the proximal femur in figure 3.6. The preferred direction coincides with the direction of principal stress, which increases bone stiffness in that direction. On top of this geometric anisotropy, the porosity and connectivity of cancellous bone also varies spatially in the femur.

In the proximal femur, the most distinct cancellous structure is the highly directional trabeculae extending from the lower aspect of the femoral neck to the top of the femoral head. Anisotropy can also be seen in the trabeculae at the distal end of the
proximal femur, where plates arch upwards, away from the inner cortical surface. In
the distal femur, trabeculae run vertically between each of the condyles and the wall
of the distal femoral shaft.

3.2 Bone Biology

Bone is very much a living tissue, consisting of mineral components continually laid-
down and resorbed by specialised cells. The activities of these cells depend on the
type of bone, mechanical forces experienced, and the presence of signalling molecules.
Figure 3.7 illustrates the microstructure of bone.

3.2.1 Composition

Bone is composed of hydroxyapatite, collagen, and water. Components of hydroxyap-
atite include calcium phosphate, calcium carbonate, sodium, magnesium, and flouride.
Hydroxyapatite forms 65% of bone by weight, and is responsible for the hardness and
stiffness of bone. Collagen forms 30% of bone by weight, and offers toughness through
the fibrillar matrix in which hydroxyapatite is embedded. Water makes up the re-
main ing 5%, and is necessary for the strengths of both hydroxyapatite and collagen
fibres.
3.2.2 Cellular Components

Bone remodelling is carried out by cells known as osteoblasts, osteoclasts and osteocytes. Osteoblasts are responsible for the laying-down of new bone (apposition), while osteoclasts remove bone (resorption). Once osteoblasts become embedded in the new bone they lay down (e.g. in cortical bone), they become osteocytes, which routinely remodels the bone around them through both bone resorption and apposition.

Cortical bone is composed of repeating units called osteons. An osteon is cylindrical in shape, and approximately 100 µm in diameter. It consists of a central lumen, known as a Haversian canal, surrounded by concentric rings of dense lamellar bone. The Haversian canal houses blood vessels, a nerve, and fluids. Small canals (canaliculi) link the Haversian canal to spaces (lacunae) in the lamellar bone where osteocytes reside. The exterior surface of cortical bone is covered by the periosteum - a thin layer of fibrous tissue. The periosteum is important for bone's blood supply, lymph drainage, fracture healing, and development. Cells in the periosteum remodel the outer layer of cortical bone.

Cancellous bone structure is porous, consisting of a lattice of plates and rods. Remodelling of the lattice occurs through the thickening, thinning, addition, and removal of trabeculae. This is performed by osteoblasts and osteoclasts on the surfaces of trabeculae.

3.2.3 Bone Remodelling

Bone remodelling occurs at discreet sites, termed bone remodelling units (Frost, 1965), in cycles of resorption and deposition. In the resorption phase, osteoclasts remove bone over a period of two to three weeks. In cortical bone, bone is tunnelled away, while in trabecular bone, a lacunae is excavated on the surface. In the subsequent apposition phase, osteoblasts lay down new bone where old bone was resorbed. The apposition of new bone lasts around 13 weeks (100 days).

Unbalancing the resorption and apposition phases lead to overall bone growth or loss. During development, fracture healing, or increased mechanical stimuli, apposition by osteoblasts is up-regulated, so more bone is formed than removed. On the other
hand, when osteoclast activity is higher than that of osteoblasts, bone loss results. This can be due to hormonal influences, or the absence of mechanical stimuli (Parfitt 1982).

The response of bone remodelling to mechanical stimuli has long been studied. For example, Wolff (1892) linked the direction of trabecular bone in the proximal femur to principal directions of stress. Today, the theory of bone remodelling is based on three key concepts:

1. Bone remodelling optimises strength and weight,
2. Bone remodelling aligns the principal direction of trabecular bone with the principal directions of stress,
3. Bone remodelling is a self-regulated response to mechanical stimuli.

Increase in mechanical loading, for example through exercise, leads to increased Bone Mineral Density (BMD) (Devine et al. 2004), thicker cortical bone, and alignment of collagen fibres in bone (Robling et al. 2006). Cortical bone under tensile stress possess collagen fibres aligned in the direction of the load, whereas under compressive stress, possess collagen fibres aligned in planes transverse to the load (Riggs et al. 1993). These changes increase the stiffness of the bone matrix for each particular mode of loading.

The removal of mechanical stimuli results in the weakening of bone and abnormal morphologies. For example, Chalmers and Ray (1962) transplanted a developing rat femur into an unloaded environment, which resulted in femurs that lacked diaphyseal curvature (bending along the shaft), had less directional trabeculae, and a widened femoral neck. Frost (1998) showed that in mature bones, the removal of mechanical load leads to bone loss. This is backed up by Baldwin et al. (1996); Collet et al. (1997), who studied astronauts in a weightless environment, whom lost up to 7% in bone mass, even with exercise. In Warner et al. (2006) mice with paralysed hind limbs exhibited loss of trabecular and cortical bone, primarily through increase in bone resorption.

The regulation of remodelling process, and the effects of mechanical stimuli provide the means to explain changes in femoral morphology. But first, we must know about how femur morphology has traditional been measured.
3.3 Traditional Measurements

Interest in femur morphology has existed for a long time. As such, a range of morphological measurements specific to the femur already exist. General measurements of femoral size and morphology are predominantly calliper-like measurements, or angles. The measurements are typically taken on a femur placed anterior or posterior side down on a flat surface, with the shaft aligned with a horizontal axis.

This section reviews the common measurements, accompanied by their averages from an analysis of a modern Australian adult population in [Hislop-Jambrich (2010)]. Table 3.1 lists these measurements, and their averages taken from that study.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>long femur length (cm)</td>
<td>47</td>
<td>43</td>
</tr>
<tr>
<td>short femur length (cm)</td>
<td>44</td>
<td>41</td>
</tr>
<tr>
<td>femoral axis length (cm)</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>epicondylar width (cm)</td>
<td>8.4</td>
<td>7.4</td>
</tr>
<tr>
<td>mid-shaft width (cm)</td>
<td>3.0</td>
<td>2.6</td>
</tr>
<tr>
<td>femoral neck width (cm)</td>
<td>3.4</td>
<td>3.1</td>
</tr>
<tr>
<td>femoral head diameter (cm)</td>
<td>5.2</td>
<td>4.5</td>
</tr>
<tr>
<td>neck-shaft angle (degrees)</td>
<td>127</td>
<td>127</td>
</tr>
<tr>
<td>anteverision angle (degrees)</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

3.3.1 Femoral Lengths

Femur length is commonly measured by the long and short femur lengths (Fig. 3.8a, 3.8b). The long femur length is the longest distance from the superior tip of the femoral head to the inferior side of the medial condyle. The short femur length is from the superior tip of the greater trochanter to the inferior side of the lateral condyle.

Another length important for the mechanics of the femur is the femoral axis length (Fig. 3.8c). This is a measure of the moment arm created by the femoral neck and head. It is the distance between the femoral neck axis’ intercepts at the surface of the femoral head, and the greater trochanter.
3.3. TRADITIONAL MEASUREMENTS

Figure 3.8: Femur lengths. Definition of the long and short femoral lengths are illustrated in figures (a) and (b), respectively. The femoral axis length is illustrated in figure (c). Reproduced from Hislop-Jambrich (2010).

3.3.2 Femoral Widths

Femoral widths are measured between the condyles (epi-condylar width), at the mid-shaft (mid-femoral width), and below the trochanters (sub-trochanteric width) (Fig. 3.9). These widths are well correlated to gender and stature. In addition, femoral neck width, measured at the narrowest part of the neck, is important for determining femur strength. Femoral head diameter is a useful determinant of gender.

3.3.3 Axes and Angles

The angle formed between the femoral neck axis, and the femoral shaft axis is the neck-shaft angle (Fig. 3.10). The neck-shaft angle is an important determinant of hip fracture risk. The anteversion angle describes the twisting of the femur, from the distal to the proximal end. It is defined as the angle between the femoral neck axis and the knee axis, in a plane perpendicular to femoral shaft axis.
3.3.4 Femoral Cortex

Cortical bone is most commonly measured in terms of its thickness. Cortical thickness is often measured at various sites, including the mid-shaft, in the sub-trochanteric region, and the superior and inferior regions of the femoral neck. In 2-D cross-section slices, periosteal and endosteal perimeters have also been used (Bertelsen et al. 1995; Mayhew et al. 2005; Holzer et al. 2009). These measurements can be made in a clinical setting, on planar or CT X-ray images. In the micro-scale, cortical bone porosity and osteon size are often used as a measure of cortical bone quality, or for predicting age (Jones et al. 2008; Bertelsen et al. 1995). These measurements are made in the mid-shaft, where cortical bone is thick, robust and easier to handle. However, due to the need for invasive high-resolution imaging, these measurements are out of reach for clinical images.
3.3. TRADITIONAL MEASUREMENTS

3.3.5 Cancellous Bone

The complexity of cancellous bone structure is matched by the number of ways it is measured. Measurements exist to characterise trabecular density, porosity, connectivity and anisotropy (Odgaard, 1997). The resolution required for these measurements are in the range of micro-CT imaging. Collections of sample are invasive, and limited to a small number of sites.

On a larger scale, the trabecular margin length (TML) (Fig. 3.11) can be used to measure the extent of cancellous bone in the proximal femur (Hislop-Jambrich, 2010). It is defined as the distance along the shaft axis, from the most distal edge of proximal femur cancellous bone to the inner cortical surface at the junction of the greater trochanter and the femoral neck.

3.3.6 Bone Mineral Density

Bone mineral density is commonly measured using Dual-energy X-ray Absorptiometry (DXA), or quantitative X-ray CT (QCT) imaging (Fig. 3.12). Both methods take
Figure 3.11: The trabecular margin length, measured from the lowest point of the femoral neck, to the highest point of the lower boundary of proximal femur cancellous bone. Reproduced from Hislop-Jambrich (2010).

Figure 3.12: DXA (a) and QCT (b) images for measuring bone mineral density. Figure (a) reproduced from Germain et al. (2010). http://dx.doi.org/10.1186/1750-1172-5-30

advantage of the fact that mineral elements of bone absorb X-ray radiation. The higher the BMD, the higher the absorption, so image intensity can be used to infer BMD.

DXA is the most widely used method to measure bone density (Hislop-Jambrich
A DXA image is produced from two X-ray images taken at different energies. The differences in absorption between the two images is used to calculate areal bone mineral density (aBMD). The areal prefix comes from the fact that DXA images are the 2-D projections of 3-D BMD variations. aBMD measurements are sensitive to variations in patient orientation, and surrounding tissue, making its accuracy questionable (Bolotin, 2007).

QCT uses computed tomography to produce 3D images, from which volumetric bone mineral density (vBMD) can be calculated. In QCT imaging, a phantom containing materials with known mineral densities is imaged with the specimen. A function mapping image intensity to mineral density can be found by regression on phantom image intensity and their known densities. Because QCT is a 3-D imaging modality, it gives a more precise measure of BMD compared to DXA. However, it is more expensive in term of time and cost, and gives a higher radiation dose. Chapter 5 will present further details on QCT, and the characteristics of the Victorian Institute of Forensic Medicine (VIFM) QCT femur images.

### 3.4 Natural Variations

In a natural population, femur morphology variations result from a combination of genetic, physiological and environmental factors. Variations exist across age groups, genders and ancestry. Bone remodelling, body stature, diet and physical activity over a lifetime introduces more sources of variation. This section reviews morphological variations related to these factors.

#### 3.4.1 Development

In the fetus, the femur consists of five ossification centres (femoral head, condyles, greater and less trochanters, femoral shaft) connected by cartilage. Bone formation expands from these centres, meeting at epiphyseal lines in femoral head, greater trochanter, and condyles, around the age of 20 (White and Folkens, 2005).

All size-related measurements increase during development, along with cortical thickness, which on average reaches 6 mm for females, and 7 mm for males, by 20
years of age (Bertelsen et al., 1995) (Fig. 3.13). In addition, anteversion decreases from between 30 and 40 degrees at birth, to 8 to 14 degrees upon adulthood (Gulan et al., 2000).

Bone remodelling in response to mechanical stimuli appears to be most sensitive during development (Welten et al., 1994; Kannus et al., 1995; Bradney et al., 1998). In this period, not only can exercise increase bone density, but the benefits can persist into adulthood. Bass et al. (1998) showed that retired gymnasts, who were active during prepubescent years, have significantly higher BMD in load bearing bones, compared to control. On the other hand, Frisch et al. (1981) showed that heavy exercise during puberty may disrupt hormone production, leading to lower peak BMD. After puberty, exercise combined with lowered body weight has shown to result in bone loss and decreased BMD (Drinkwater et al., 1984). In summary, the load-history during development has a major influence on femoral morphology. Conversely, adult femoral morphology can be used to infer physical activity levels earlier in life.

### 3.4.2 Ageing

Ageing is generally associated with a gradual loss of bone mass. Morphologically, the symptoms are thinning of cortical bone, increase in cancellous bone porosity, and decrease in trabecular connectivity. Along with the decrease in bone density, these morphological changes weaken the femur, leading to a higher risk of fracture. Bone loss results from increased bone resorption relative to apposition, which is correlated to age-related decrease in oestrogen and testosterone levels, for females and males respectively (Parfitt, 1982). The lack of load-bearing activities in old age, due to lifestyle changes and muscle wasting, also reduces the formation of new bone (Vainionpää et al., 2007).

In the femoral mid-shaft, cortical bone variation with age was studied in detail by Bertelsen et al. (1995). The authors found that from around the age of 20, cortical thickness decreased gradually, along with cross-sectional area (Fig. 3.13). The rate at which it decreased was higher in females than in males, especially over the age of 50. Meanwhile, at least for males, periosteal perimeter continued to increase beyond the age of 20 at a much lower rate. However, this last observation was not seen by other researchers (Ruff and Hayes, 1982; Cohen, 1990). Overall, there is an age-
related decrease in cortical thickness, with slight or no increases in periosteal perimeter, particularly in females. This suggests that cortical bone loss with age is due to increased bone resorption on the endosteal surface. In the femoral neck, for an adult population, cortical thinning and expansion also occur, although primarily the superior and inferior regions (Carpenter et al. 2011).

Age-related bone loss also impacts cancellous bone. In the proximal femur, unpublished results from Hislop-Jambrich (2010) showed that cancellous bone recedes proximally with age. In a comparison between males below and above 50 years of age, trabecular margin length was 15% longer in the former group. A difference of 12% was reported for females. Bone mineral density steadily declines with age. In men, the rate of decrease is around 1 g/cm³ until around the age of 70, when it accelerates. In women, the acceleration occurs early, at about the age of 50, coinciding with menopause (Jones et al. 2008).
3.4.3 Gender

As listed in table 3.1, most morphological measurements of the femur are greater for male than females. The ratio between femoral lengths and widths suggests that male femurs tend to be thicker or broader. Male femurs possess thicker cortical bone and significantly higher BMD (Jones et al., 2008). Larger male femurs reflect larger male stature, weight and musculature in general. The size difference is present even after adjusting for body size difference (Looker et al., 2001).

Age-related bone loss is more severe in females than males (Fig. 3.14), mainly due to hormonal changes after menopause. This, combined with thinner cortical bone, results in female femurs being more susceptible to fracture with increasing age. Reproduction introduces another source of variation in females, particularly in bone quality. Preg-
3.4. NATURAL VARIATIONS

3.4.4 Ancestry

The ancestry of a person, in terms of geographical origin, can be estimated from skeletal morphology (Brace, 1995). While not as reliable as the skull, a number of discriminatory features have been identified on the femur. For example, the anterior-posterior diameter of the proximal femur has been found to be larger in whites and blacks than in East Asians and American Indians (Gill, 2001). By combining a number of measurements, discriminatory powers can be increased, as shown by Ballard et al. (1999) using 13 femoral measurements to discriminate between black and white individuals.

Racial variations in femur morphology has also been investigated in an attempt to explain variations in hip fracture risk. Nakamura et al. (1994) found shorter femoral necks and smaller neck-shaft angles amongst Japanese women, compared to North American whites. This was used to explain Japanese womens’ lower risk of femoral neck fracture, despite having lower femoral neck bone mass. In an analysis of North American male proximal femurs, Travison et al. (2008) found higher BMD, cross sectional area, and section modulus in African Americans, while age-related loss in BMD was greater in Hispanics. The evidence above suggests that there are ancestry-related variations in femur morphology. However, establishing the relationship to femur strength and fracture risk is still a focus of ongoing work, and can definitely benefit from larger sample sizes and higher fidelity measurements.

3.4.5 Stature, Body Mass and Other Factors

Femur length is a reliable predictor body height, or stature. The relationship has been quantified in many studies for differing populations (Formicola, 1993). As an example, Trotter (1970) found the linear relationship between stature and the long femoral length

BMD decreases during pregnancy (Poulsen et al., 2001), but recovers in the long term (Pearson et al., 2004). In addition, there is evidence that parity (the number of children born) increases BMD (Murphy et al., 1994).
in white males to be

\[ \text{stature(cm) } \pm 3.27 = 2.38 \times \text{femur length} + 61.41 \]  

(3.1)

In more complicated methods, dimensions such as epicondylar width are also used (Oliver [1976]).

Greater body mass imposes greater downward force on the femur at the hip joint. A larger femoral head will spread the force over a greater area, reducing stresses and strains in the femur. Therefore, it is not surprising that body mass is correlated to femoral head diameter. Auerbach and Ruff (2004) compared a number of methods for estimating human body mass from femoral head diameter, and found good correspondence between methods.

Higher body mass can be correlated to larger muscles and higher muscle forces (Jones et al. 2008). Higher loads all around results in a stronger femur morphology. Ruff (2003) showed that femur strength was positively correlated with body weight and thigh muscle size in adolescents. The increase in femur strength was in the form of increased cross-sectional modulus, meaning increased cortical thickness and femoral shaft width. Higher physical activity is also positively correlated with these measures (Bridges 1991). However, the effect of physical activity on other aspect of femur geometry, and cancellous bone structure, is less well known (Jones et al. 2008).

Lastly, nutrition plays a life-long role in bone strength. Devine et al. (2004) showed that high dietary calcium intake with high levels of physical activity improves hip BMD. Conversely, Marcus and Menczel (2007) showed that poor diet leads to more severe bone loss later in life.

3.5 Pathologies

Femur morphology can be a useful indicator of two common femur pathologies: osteoporosis (OP) and osteoarthritis (OA). It can also be used to estimate hip-fracture risk. With ageing populations world-wide, the prevalence of OP, OA and hip-fracture will increase along with the burden on public health systems. Early diagnosis and drug-treatment can mean better treatment outcomes. But perhaps more importantly,
understanding how certain morphologies cause dysfunctions, or vice-versa, can lead to better prevention or management strategies.

### 3.5.1 Osteoporosis

![CT images of an osteoporotic proximal femur (a) and a normal case (b).](image)

**Figure 3.15**: CT images of an osteoporotic proximal femur (a) and a normal case (b). The osteoporotic case has noticeably lower BMD in its cancellous bone.

According to the World Health Organisation (WHO), osteoporosis is defined by bone mineral density 2.5 standard deviation below the mean peak value of a healthy, young adult, as measured by DXA (see Sec. 3.3.5). Osteoporosis is widespread, affecting over 75 million people in Europe, United States, and Japan alone. Osteoporosis is classified by its age of onset. Primary type 1 osteoporosis is associated with menopause in woman. Primary type 2 describes osteoporosis occurring beyond the age of 75, which affects twice as many women as men. Secondary osteoporosis affects both genders equally, and can begin at any age (Hislop-Jambrich 2010).

Osteoporosis has the most obvious effect on cancellous bone structure (Fig. 3.15). This is because cancellous bone has more surface area than cortical bone on which
Figure 3.16: CT image of an osteoarthritic hip joint. Cartilage degeneration has left little or no space between the femoral head and the pelvis.

over-active bone resorption takes place (Seeman and Delmas 2006). Symptoms include the breakage and loss of trabecular struts and rods, and consequently a decreases in trabecular connectivity and bone strength. As a result, osteoporotic fractures tend to occur in regions that rely on cancellous bone for strength. Examples include the proximal femur, wrist, and spine. The International Osteoporosis Foundation estimates that in people over the age of 50, one in three women, and one in five men, will suffer a osteoporosis-related fracture.

3.5.2 Osteoarthritis

Osteoarthritis is characterised by the degeneration of cartilage and subchondral bone (Grynps et al. 1991). According to Pinto et al. (2011), 14.8% of New Zealand adults are diagnosed as having OA, with the figure predicted to rise. Symptoms of OA include pain, loss of mobility, the narrowing of joint separation (Fig. 3.16), excess bone formation around joints (osteophytes), and in some cases, joint effusions. The causes of osteoarthritis include excessive or abnormal loading at joints, age-related degeneration of cartilage, and genetic factors (Carter et al. 2004). Bone-remodelling, abnormal morphology and abnormal loading are all interrelated, and they influence the onset and progression of OA. Understanding variations in femur morphology could inform the diagnosis, prevention or treatment strategies for osteoarthritis, at the hip and knee.
joints. In cases where joint replacement is necessary, the information can be used to select the most appropriate prosthesis.

### 3.5.3 Proximal Femur Fractures

It is believed that proximal femur fractures (Fig. 3.17) contribute to a major proportion of deaths due to diseases of muscles, bones and tendons ([Kanis et al., 2003; Abrahamsen et al., 2009](#)). In addition, immobility following a fracture severely compromises the quality of life for patients, and increases the burden on caregivers. Proximal femur fractures can be categorised into four types based on location ([Hislop-Jambrich, 2010](#)): sub-capita, transcervical (neck), inter-trochanteric, and sub-trochanteric. Fractures occur when the femur is loaded beyond its fracture stress. This fracture stress is dependent on the magnitude of loading, the direction and location of loading, spatially varying bone material properties, and the geometry of the femur itself. The increasing prevalence of hip fracture, in ageing populations, means much effort has been put into quantifying the relationships between these factors and fracture risk.

Of course, osteoporosis is an important factor that affects bone strength, and therefore fracture risk (Sec. 3.5.1). Higher rates of osteoporosis in women and the elderly
correlate to higher risks of hip-fracture. Bone loss through other means, such as the absence of mechanical stimulus due to immobility (Mikkola et al., 2007) also increase the risk. However, geometric factors also come into play (Kaptoge et al., 2008). For example, Nakamura et al. (1994) found a relationship between larger neck-shaft angles in Caucasian women, and their higher fracture risk, compared to Asian or African American women. The offered explanation is that a higher neck-shaft angle leads to lower bending moments in the neck, and hence thinner cortical bone. Upon a side-impact, such as a fall, high bending and shear forces are experienced in the relatively weak neck, causing fracture. Following this reasoning, it is unsurprising that high neck-shaft angle can also be used to predict transcervical fractures (Partanen et al., 2001).

Cortical thickness plays an important role in fracture risk. Being roughly cylindrical in shape, cortical thickness is the main determinant of the femur’s bending stiffness. Recent works by Holzer et al. (2009); Verhulp et al. (2008) have suggested that cortical thickness distribution is more important than cancellous bone structure in fracture resistance. Past studies such as Partanen et al. (2001) and Gregory and Aspden (2008) surveyed cortical thickness at few individual sites. Recently, works such as Treece et al. (2010, 2012) introduced methods for measuring cortical thickness distribution over the whole femur, which can lead to better cortical thickness-based fracture predictions (Li et al., 2009).

3.6 Statistical Modelling of Femur Morphology

Instead of using traditional morphology measurements, research from the field of statistical shape analysis (SSA) (see chapter 2) have tried to incorporate continuous 3-D femur shape and appearance information into population analysis. SSA has been used for three main applications: image segmentation, model generation, and classification and prediction. This has lead to more accurate and sensitive femur models for finite-element analysis or computer-guided surgery, and predictors for fracture risk. This section will review recent works in the aforementioned three applications.
3.6. STATISTICAL MODELLING OF FEMUR MORPHOLOGY

3.6.1 Segmentation

Statistical shape models have been widely used for model-driven femur segmentation. The shape model constrains the segmentation to realistic shapes, making it robust against noise or faint boundaries. Femur segmentation from X-ray CT or MRI images has been mainly driven by the needs of computed-guided surgery. Since hip and knee arthroplasties are the two major femur surgical procedures, segmentation has focused on the hip and knee joints.

Many statistical shape model-driven segmentation methods have been developed, for different image modalities, organs, and making use of different image information. Heimann and Meinzer (2009) provides a thorough review of recent literature. One method in particular, Active Shape Modelling (ASM) (Cootes et al., 1994), is widely used for the femur. ASM combines statistical models of the image normal to the object boundary, with a statistical model of object shape, for robust segmentation (chapter 7 and its appendices present the theory and implementation of ASM in detail).
Josephson et al. (2005) used ASM to segment the distal femur from 3D MRI knee images. The authors reported that successful segmentation was sensitive to the initial position and pose of the generated model. Heinze et al. (2002) used an ASM framework to segment all the bones of the knee joint. The femur was segmented first, with manual initialization. The initial position and pose of the other bones were then inferred from the segmented femur. However, validation of the method’s accuracy was not provided. In Fripp et al. (2007), initialization was robust and automatic, using affine registration of a reference image. In addition, after ASM segmentation, points on the generated model were fitted to the bone-cartilage interface, free of shape model constraints, to obtain a more accurate segmentation (Fig. 3.18). In leave-one-out validation experiments, surface overlap of 94% was achieved at the interface. Ramme et al. (2011) presented an interesting approach where probability maps of bone boundary of the knee were produced on a training set of registered images. The model image was then deformed to match data knee images using expectation maximisation, thereby segmenting the bone surface geometry.
3.6. STATISTICAL MODELLING OF FEMUR MORPHOLOGY  

In the hip, Lamecker et al. (2004) showed that manually initialised ASM segmentation can achieve a root mean squared (RMS) error of 2.4 mm for the pelvis, compared to manual segmentation. In Seim et al. (2008), an RMS error of 1.9 mm was achieved. Schmid et al. (2011) presented a method for segmenting both the pelvis and femur from images with limited field of view (Fig. 3.19). Their goal was to predict the full geometries of the pelvis and femur from images cropped to the hip joint. The segmentation was initialised by manual definition of the hip-joint centre (HJC). During segmentation, the generated pelvis and femur models were constrained to rotate and translate (to a small degree) about this centre. Also, during segmentation, the ASM was only allowed to use local appearance models that lay within the field of view of the image. Errors of about 2.0 mm RMS were reported.

The proximal and distal portions of the femur are the most challenging for automatic segmentation. Close neighbouring bones, and low bone density mean that contrast at the femoral surface is low. However, as the works above have shown, ASM segmentation can be robust in these environments.

3.6.2 Model Generation

Statistical models of shape and morphology are generative models. For example, new femurs can be generated from a linear combination of the principal components. This provides a way to simulate a large sample size for population-based studies. Alternatively, femurs can be generated to match sparse morphological data, and other information such as age, body weight and height.

Querol et al. (2006) generate femurs to study variations in mechanical behaviour associated with shape and BMD variations. A statistical deformation model of quantitative-CT images was used to capture both femur shape and BMD distribution. Finite element analysis was performed on femurs reconstructed along principal components to reveal internal stresses and strain variations that correlated with morphological variations.

A simulated large-sample size study was done in Bryan et al. (2009), where a statistical model of femoral shape and bone density distribution (Bryan et al. 2010) was used to randomly generate 1000 femurs. Finite element analysis (FEA) on these femurs was used to determined those likely to fracture (Fig. 3.20). The morphology of
these were then compared to femurs unlikely to fracture. The study identified proximal femur cortical thickness as a key difference, as expected. Then in Bryan et al. (2012), the same approach was used to investigate the effects of morphological variations on hip implants. A limitation of this series of work is that the statistical model was trained on only 21 femurs, which may not represent the full range of natural variations.

Indeed, FEA is commonly performed on femur geometries to predict fracture location, or evaluate prosthesis performance. However, FEA is computationally expensive. Using synthetic shapes, Khalaji et al. (2008) demonstrated that a statistical model can be trained on both shape and pre-computed stress and strain information. The model can then be used to efficiently approximate FEA results based on shape information.

The generation of femurs can be guided by sparse or complete data. For example, given a small set of target landmarks, principal component weights can be optimised to generate a femur with landmarks that best matches the target landmarks. This is an efficient way to generate 3D models for surgery planning, or generating patient-specific models without expensive 3-D imaging and segmentation. Landmarks collected during
3.6. STATISTICAL MODELLING OF FEMUR MORPHOLOGY

Figure 3.21: Statistical model reconstruction of femur geometry from ultrasound scans. Reproduced from Barratt et al. (2008), http://dx.doi.org/10.1016/j.media.2007.12.006

Surgery were used to reconstruct knee and femur surface models in Fleute et al. (1999); Stindel et al. (2002); Rajamani et al. (2005) and Rajamani et al. (2007). Surface models of the femur and pelvis were reconstructed from partial ultrasound scans in Barratt et al. (2008) (Fig. 3.21). Reconstruction accuracy is usually improved with more target data. However, that target data can also be in the form of surrogate measurements such as body weight, height, age and gender - data that is correlated with femur morphology (Blanc et al. 2012).

Sparse data can also come in the form of one or more 2-D images, such as planar X-rays. Compared to X-ray CT, planar X-ray is less hazardous in terms of radiation dosage, and more economical. These are good reasons for reconstructing 3D geometries from 2D images. This was demonstrated on the distal femur in Fleute and Lavallee (1999). A 3-D femur model was generated to minimise the error between the femur boundary in the planar image, and the silhouette of the reconstruction. Dong and Zheng (2009) used a simplified 2D geometric model of the proximal femur to better align the shape model prior to model generation. Proximal femur reconstruction from
DXA images was presented in Whitmarsh et al. (2010). In this case, the model was of both shape and CT image appearance. The reconstruction minimised the error between the DXA image and simulated DXA images produced from the reconstructed femur.

3.6.3 Clinical Prediction

In the femur, classification studies are focused on identifying osteoporosis, predicting fracture risk, and finding correlations between morphology and osteoporosis or osteoarthritis. In general, SSA using 3-D femur models have produce equal or better results compared to traditional femur measurements.

![Figure 3.22: Low and high fracture risk morphologies predicted from a statistical model of the proximal femur. Reproduced from Whitmarsh et al. (2011), http://dx.doi.org/10.1007/978-3-642-23629-7_48](image)

Attempts to predict hip fracture risk from femoral morphology has a long history, particularly using traditional one dimensional measurements, such as lengths and angles. Statistical models of shape and morphology has been shown to improve the accuracy of predictions (Gregory and Aspden, 2008). For example, Gregory et al. (2004) combined a 2D shape model of the proximal femur with BMD measurements from five sites for discriminant analysis between fracture and non-fracture groups. By combining shape and bone density, the author produced a fracture classifier with an accuracy of 90% on a limited data set. In Whitmarsh et al. (2011), a 3-D model of proximal shape and BMD distribution was used predict fracture risk using the same statistical meth-
3.7 Limitations of Previous Works

It is clear that population-based morphological analysis of the femur has benefited from statistical shape modelling techniques. However, there are four main limitations with the previous works reviewed:

1. The population size of previous works is often limited, especially for the whole femur;
2. There is a lack of a complete system for image segmentation, morphology modelling, and statistical analysis, which would streamline data collection;
3. There does not exist a femur description that recognises cortical and cancellous bones as separate regions;
4. All previous work for the femur have used dense linear meshes, or registered images to model the femur, which results in a very large number of parameters for statistical analysis.

Limited training set size limits the power of any statistical model to generalise a population, which is one of the main purposes of a statistical shape model. In the works reviewed, 235 was the largest training set, used for analysis of the knee by Heinze et al. (2002), whereas most other used a training set of less than 100. In the VIFM dataset used for this work, there are 262 images.

To process such a number of images, and possibly many more in the future, an automated pipeline for segmenting, modelling, and analysing whole-femur morphology is required. The fundamental methods required for the pipeline exist in various implementations, for various applications. However, there lacks a unified system for the whole process. Segmentation methods for the femur (using ASM) tend to ignore internal features, such as the inner cortical surface and cancellous bone, while morphology modelling and analysis works tend to use pre-segmented images (Bryan et al., 2010; Whitmarsh et al., 2011). In addition, no previous work has treated cortical and cancellous bone as two separate regions. Separating the regions gives the freedom of analysing cortical geometry and cancellous BMD distribution independently, or together as needed. As an example, cortical thickness distribution and variation can be analysed by isolating cortical geometry, while BMD distribution in cancellous bone can be studied without the confounding effects of the much denser cortical bone.

The way in which the femur is modelled has either been a dense linear mesh (for point-distribution models), or registered 3-D images (for statistical deformation models). While these descriptions are simple and convenient, their resolution must be high to be of acceptable accuracy, and so result in a very large number of parameters for statistical analysis. For example, the whole-femur volumetric mesh in Bryan et al. (2010) was composed of 117,225 vertices, each carrying 4 parameters (3-D coordinates
plus BMD), giving a total of 468,900 parameters for principal component analysis. Using higher order descriptions, such as piecewise parametric meshes, can reduce the number of parameters significantly, reducing computational cost and producing more compact statistical models.

Therefore, the goal of this thesis is to build an automated pipeline for segmenting, compactly modelling, and analysing whole-femur morphology, from a large population of CT images. Chapter 4 presents the general theories and methods required for the pipeline, while Chapter 5 presents the data set used. Chapter 6 describes the design and training of a femur shape model, based on relatively compact high-order piecewise parametric meshes. Chapter 7 then describes the automatic cortex segmentation and modelling method, which is driven by the shape model, while chapter 8 describes cancellous bone image segmentation and registration.
Chapter 4

General Theory

This chapter serves as a reference section for the theory behind key methods used in the CT-to-model pipeline, providing details and examples required for their implementation. The pipeline flowchart is repeated here in figure 4.1 for convenience.

Covered topics are:

- Principal Component Analysis (PCA), for training the femur statistical shape model;
- Lagrange piecewise parametric meshes, used to describe femur shape;
- Mesh fitting, used in femur shape model training, and modelling the cortical surfaces;
- Active Shape Modelling (ASM), adapted for the Lagrange mesh for automatic femur segmentation;
- Cortical Thickness Mapping (CTM), adapted for the Lagrange mesh for segmenting the inner and outer cortical surfaces and measuring cortical thickness;
- Radial Basis Functions (RBFs), for registering image volumes of cancellous bone.

The use and integration of these methods into the pipeline will be described in the subsequent chapters. Relevant sections in this chapter will be frequently referred to for readers looking for more details. The subsequent chapters will also discuss the reasons for using these methods, their advantages, and their disadvantages in terms of the pipeline.
Figure 4.1: Main steps and methods of the automatic femur CT-to-model pipeline.
4.1 Principal Component Analysis

Principal Component Analysis (PCA) is the most widely used dimension reduction method for statistical shape analysis, and is used throughout the femur CT-to-model pipeline. PCA decomposes multi-dimensional data into orthogonal components of maximum variance. Given \( n \) observations of \( m \) variables, or \( n \) data points in \( m \)-dimensional space, PCA can be thought of as a rotation of axes in \( m \)-dimensional space so that the first axis is aligned with the direction of maximum variance amongst the \( n \) points. The second axis will be aligned with the next greatest direction of variance that is orthogonal to the first, and so on. This is illustrated in figure 4.2 on 2-D data. The directions of these new axes are the principal components, and the projections of the data points on the components are referred to as scores, loadings, or weights. Since principal components are orthogonal, observation can be reconstructed by a linear combination of principal components weighted by its scores, or closely approximated by the first few principal components and associated scores. PCA assumes that the data is Gaussian (can be described by the mean and variance), and that the data has a high signal-to-noise ratio so that high variance has significance.

4.1.1 Computation

There are two general ways to perform PCA: eigen-decomposition of the covariance matrix, and singular value composition (SVD) (Golub and Reinsch, 1970). In both cases, the \( n \) observations of each of the \( m \) variables are mean-centred, and concatenated into an \( m \) by \( n \) data matrix \( X \)

\[
X = \begin{pmatrix}
    x_{1,1} & \cdots & x_{1,n} \\
    \vdots & \ddots & \vdots \\
    x_{m,1} & \cdots & x_{m,n}
\end{pmatrix}
\]  

(4.1)

In the eigen-decomposition method, the principal components are found by diagonalising the covariance matrix of \( X \), \( C_X \)

\[
C_X = \frac{1}{n-1}XX^T
\]

\[
= EDE^T
\]
The eigenvectors of $\mathbf{C}_\mathbf{X}$ in the columns of $\mathbf{E}$ are the principal components. The eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_n$, in the diagonals of $\mathbf{D}$, are the variances along each component.

When $m$ is greater than $n$, $\mathbf{C}_\mathbf{X}$ is rank deficient. There can only be as many principal components as the number of observations minus one. SVD can be used in this case, and also for greater computational efficiency (Shlens 2005). A different data matrix $\mathbf{Y}$ is constructed as

$$\mathbf{Y} = \frac{1}{\sqrt{n-1}} \mathbf{X}^T \quad (4.2)$$

such that $\mathbf{Y}^T \mathbf{Y} = \mathbf{C}_\mathbf{X}$. Carrying out SVD on $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{U} \Sigma \mathbf{V}^T \quad (4.3)$$

yields the principal components in the columns of $\mathbf{V}$, and the square root of their variances in the diagonals of $\Sigma$. 

**Figure 4.2:** PCA on 2-D data. Principal components are denoted by the red lines, with the longest being the first principal component. The length of the lines denote the calculated standard deviation in each direction.
4.1. PRINCIPAL COMPONENT ANALYSIS

Figure 4.3: Reconstruction of 2-D data using the first principal component.

The scores $\mathbf{a}$ of an observation $\mathbf{x}$ on principal components $\mathbf{b}_1, \ldots, \mathbf{b}_{n-1}$ can be calculated by

$$
\mathbf{a} = \mathbf{x} \cdot \begin{pmatrix}
\vdots & \vdots \\
\mathbf{b}_1 & \mathbf{b}_n \\
\vdots & \vdots 
\end{pmatrix} \quad (4.4)
$$

Conversely, an observation $\mathbf{x}'$ can be approximated by a linear combination of the first $k$ principal components and the mean $\overline{\mathbf{x}}$

$$
\mathbf{x}' = \overline{\mathbf{x}} + \sum_{i=1}^{k} \mathbf{b}_i a_i \quad (4.5)
$$

This is illustrated in figure 4.3, where the 2-D data from figure 4.2 is reconstructed using just the first principal component, resulting in the removal of noise in the $x = -y$ direction.

The distance of an observation from the training set mean is the Mahalanobis
distance

\[ h = \sqrt{\sum_{i=1}^{k} \frac{a_i^2}{\lambda_i}} \] (4.6)

\( h \) quantifies the “unlikeliness” of an observation, and can be used to identify outliers, or penalise the generation of unrealistic observations.

The number of components \( k \) is usually chosen so that the variance accounted for \( \sum_{i=1}^{k} \lambda_i \) reaches a certain proportion of total variance \( \sum_{i=1}^{n-1} \lambda_i \). The proportion is usually arbitrarily chosen at 0.95. More objective methods for determining \( k \) include find the “elbow” in the PCA scree plot (Jolliffe 2002), parallel analysis (Horn 1965), and cross-validation (Wold 1978). Another method is to compare PCA performed on incrementally more data, to PCA performed on noise to determine the optimal training set size and number of modes (Mei et al. 2008).
4.2 Piecewise Parametric Meshes

A piece-wise parametric surface mesh is an ensemble of patches, or 2-D elements, each interpolated by polynomial basis functions (Fig. 4.4). The mesh is a natural and flexible way to describe surfaces, and is used to model femur shape in the femur CT-to-model pipeline. Piece-wise parametric meshes offer a number of advantages over triangulated surfaces for statistical shape modelling:

- High-order polynomial parametric meshes typically require fewer degrees of freedom to represent complex surfaces. This reduces the size of the correspondence problem (correspondence needs to be ensured on fewer degrees of freedom), and the complexity of the subsequent statistical model;
- A triangulated surface has a fixed number of vertices on which image analysis can be performed. In contrast, parametric meshes can be discretised to any desired
resolution;
• The basis function of mesh elements constrain the way they fit to certain geometries, which helps preserve correspondence when fitting meshes to a population of shapes;
• The smooth nature of parametric meshes makes them more robust when representing noisy data.

For each element in a mesh, the basis functions $\phi_1, \ldots, \phi_p$ are defined over the element’s 2-D parametric space, with element coordinates $(\xi_1, \xi_2)$ such that

$$x = \sum_{i=1}^{p} x_i \phi_i(\xi_1, \xi_2) \quad (4.7)$$

$x = (x_1, x_2, x_3)$ are the coordinates of the mesh in $\mathbb{R}^3$, at $(\xi_1, \xi_2)$. $x_1, \ldots, x_p$ are known $\mathbb{R}^3$ surface coordinates at fixed element coordinates, called nodes or control points.

When elements are joined to form a mesh, adjacent elements share the coordinates of their nodes on their common boundary. This creates a continuous surface across elements, as shown on the mesh in figure 4.4. A point on the mesh can be uniquely defined by a set of element coordinates and the element it is in. Such a point will be referred to as a *material point*, as it is defined in the coordinates of the material, i.e. the surface. The term *mesh parameters* shall refer to the set of all nodal coordinates of a mesh. Mesh parameters $X$ is defined as:

$$X = (x_1, \ldots, x_n) \quad (4.8)$$

$$x_i = (x_i^1, x_i^2, x_i^3) \quad (4.9)$$

where the superscript in equation 4.9 denotes the coordinate direction in $\mathbb{R}^3$.

The 3-D geometry of the mesh can be altered by changing its mesh parameters (nodal coordinates). Altering mesh parameters to fit a mesh to a data cloud digitised from a surface is the focus of section 4.3. No derivative information is carried at nodes of Lagrange elements, so there is no inherent smoothness across elements, making Lagrange meshes $C^0$ continuous. Methods for enforcing smoothness during fitting is presented in section 4.3.2.
4.2. PIECEWISE PARAMETRIC MESHES

4.2.1 Quartic Lagrange Basis Function

In chapter 6, meshes with quartic (4th order) Lagrange polynomial basis functions are used to model the shape of the femur, and fitted to segmented femur geometry. A combination of quadrilateral and triangular elements are used to allow greater flexibility in mesh design, as the topology of the mesh is not constrained to a regular grid of quadrilateral elements. Also, since Lagrange elements do not have derivative parameters, there is no requirement for continuous element coordinate directions across elements.

![Figure 4.5: Quartic Lagrange line element. Node numbers correspond to basis functions in equation 4.10.](image)

The basis functions for a 1-D quartic Lagrange element, corresponding to nodes in figure 4.5 are:

\[
\begin{align*}
\phi_1 &= \frac{1}{3} (32\xi^4 - 80\xi^3 + 70\xi^2 + 3) \\
\phi_2 &= \frac{1}{3} (-128\xi^4 + 288\xi^3 - 208\xi^2 + 48\xi) \\
\phi_3 &= \frac{1}{3} (192\xi^4 - 384\xi^3 + 228\xi^2 - 36\xi) \\
\phi_4 &= \frac{1}{3} (-128\xi^4 + 224\xi^3 - 112\xi^2 + 16\xi) \\
\phi_5 &= \frac{1}{3} (32\xi^4 - 48\xi^3 + 22\xi^2 - 3\xi)
\end{align*}
\]

(4.10)

The tensor product of Eq. 4.10 gives 25 2-D element basis functions, corresponding to nodes in figure 4.6:

\[
\phi_{ij} = \phi_i(\xi_1) \phi_j(\xi_2) \quad \text{for } i, j \in \{1, 2, 3, 4, 5\}
\]

(4.11)

For a triangular element, the basis functions corresponding to nodes in figure 4.7.
Figure 4.6: Quartic Lagrange 2-D quadrilateral element. Node numbers correspond to basis functions in equation 4.11.

Figure 4.7: Quartic Lagrange 2-D triangular element. Node numbers correspond to basis functions in equation 4.12.

are
\[
\phi_{11} = \frac{32}{3} (l_1 - 3/4) (l_1 - 1/2) (l_1 - 1/4) l_1 \\
\phi_{12} = \frac{128}{3} (l_1 - 1/2) (l_1 - 1/4) l_1 l_2 \\
\phi_{13} = 64 (l_1 - 1/4) l_1 (l_2 - 1/4) l_2 \\
\phi_{14} = \frac{128}{3} (l_2 - 1/2) (l_2 - 1/4) l_1 l_2 \\
\phi_{15} = \frac{32}{3} (l_2 - 3/4) (l_2 - 1/2) (l_2 - 1/4) l_2 \\
\phi_{21} = \frac{128}{3} (l_1 - 1/2) (l_1 - 1/4) l_1 l_3 \\
\phi_{22} = 128 (l_1 - 1/4) l_1 l_2 l_3 \\
\phi_{23} = 128 (l_2 - 1/4) l_2 l_3 \\
\phi_{31} = 64 (l_1 - 1/4) l_1 (l_3 - 1/4) l_3 \\
\phi_{32} = 128 (l_3 - 1/4) l_1 l_2 l_3 \\
\phi_{33} = 64 (l_2 - 1/4) l_2 (l_3 - 1/4) l_3 \\
\phi_{34} = \frac{128}{3} (l_3 - 1/2) (l_3 - 1/4) l_1 l_3 \\
\phi_{35} = \frac{32}{3} (l_3 - 3/4) (l_3 - 1/2) (l_3 - 1/4) l_3
\]
(l_1, l_2, l_3) = (1 - \xi_1 - \xi_2, \xi_1, \xi_2) are the area coordinates defined over a standard triangle with vertices [(0, 0), (1, 0), (0, 1)] in element coordinates.

Figure 4.8: Flexibility of quartic Lagrange elements. (a) illustrates the basis function’s two points of inflection on a 1-D line element. This means that a 2-D elements (b) can efficiently model protrusions and ridges, which are common on a bone surface.

Each element has two points of inflection in each element-coordinate direction (Fig. 4.8). This means a single 2-D element can represent ridge or valley-like features, protrusions, and depressions - all of which are common on the femoral surface. Since an element can accurately represent such features, they are guided to these features by the mesh fitting process, which improves the correspondence of elements between different fitted meshes. In designing the femur mesh topology, elements are placed over high-curvature features while element boundaries are placed in regions of low curvature, when possible.

4.2.2 Comparison to PDMs and Other Parametric Meshes

A parametric mesh shape description offers a number of advantages over the popular point distribution model (PDM). The mesh contains far fewer parameters (or degrees of freedom) than a typical PDM, which contain thousands to tens of thousands of parameters. Fewer parameters lead to computational saving in mesh fitting procedures, as well as in statistical analysis. A PDM has a fixed number of points on which image analysis can be performed (e.g. for image segmentation). In contrast, parametric
surfaces can be discretised to any desired resolution. Lastly, the piecewise smooth nature of the mesh makes it more robust when representing noisy data.

A disadvantage of Lagrange basis functions is that meshes are not smooth across elements. That is, there is only $C^0$ continuity across elements, and additional regularisation is required during mesh fitting to encourage smoothness (see Sec. 4.3.2). However, $C^0$-only continuity allows rectangular and triangular elements to be connected simply by sharing boundary nodes, and any number of elements to share a common vertex, leading to more flexible mesh design. Commonly used piecewise parametric basis functions, such as Hermite, Bezier, and B-splines, do possess $C^1$ or higher continuity, but require complicated subdivision or blending operations to ensure continuity in the two situations above. An attractive alternative is the triangular B-spline, also known as DMS Splines [He and Qin, 2004]. Triangular B-splines offer the topological flexibility of triangular elements, with $C^{n-1}$ continuity given an $n^{th}$ order spline. The drawbacks are that in common implementations, the surface to be meshed must be parameterised beforehand for the placement of spline knots, and each element has non-local support so that deforming an element will change mesh geometry elsewhere.

Another advantage of Lagrange elements is that mesh parameters are simply node coordinates in the same euclidean space, which better conforms with the assumptions of statistical methods such as PCA (Sec. 4.1). In contrast, Hermite meshes are governed by both coordinates and derivative values, while splines are governed by control point coordinates, and knot coordinates in parametric space. In summary, the choice of Lagrange basis functions is based on their flexibility and simplicity in mesh design and compatibility with statistical analysis.
4.3 Mesh Fitting

The geometry of a mesh can be customised by fitting to data points, e.g. vertices from a dense polygon surface, or data points digitised from 3-D scanning. In the femur CT-to-model pipeline, mesh fitting is used in shape model training (Cha. 6), and segmentation of the cortical surfaces (Cha. 7). Fitting involves optimising nodal parameters $X$ to minimise a weighted combination of data error $\epsilon_d$ and smoothing errors $\epsilon_s$:

$$\epsilon(X) = \epsilon_d(X) + w_s \epsilon_s(X)$$  \hspace{1cm} (4.13)

Minimising $\epsilon_d$ gives a good fit in terms of closeness to data, but does not guarantee a mesh without defects such as self-intersections or sharp creases. Thus $\epsilon_s$ is introduced to regularise, or smooth the mesh. This section will present formulations of $\epsilon_d(X)$ and $\epsilon_s(X)$, and the ways in which mesh nodes are optimised. These method are used for mesh fitting in the following chapters.

4.3.1 Data Error

The distance between data points and the mesh surface $\epsilon_d$ can be defined in a number of ways. Three common methods are described below.

Data Projection

![Figure 4.9: Fitting error defined by the distance between each data point and its orthogonal projection on to the mesh, which yields its closest approach.](image)
The first method finds the orthogonal projections $p_i$ of each data point $d_i$ on the mesh (Fig. 4.9) so that

$$
\epsilon_{DP} = \sum_{i=1}^{D} \|d_i - p_i\|^2
$$

(4.14)

The projection gives the closest distances. When the mesh is composed of non-linear elements (as in the case of quartic Lagrange elements), a non-linear search is required for each data point, which makes this method computationally expensive. Also, since all distances from all data points are involved, outliers can adversely influence the fit.

Nearest Material Point

![Figure 4.10: Fitting error defined by the distance between each data point and its closest material point sampled on the mesh.](image)

The second method is a nearest material point search. The mesh is discretised by evaluating at a fixed set of material points $M = \{m_1, \ldots, m_M\}$, and the closest material point $m_i$ for each data point $d_i$ is found (Fig 4.10) so that

$$
\epsilon_{DM} = \sum_{i=1}^{D} \|d_i - \Omega(m_i)\|^2
$$

(4.15)

$\Omega(m_i)$ denotes the mesh $\Omega$ evaluated at material point $m_i$. The evaluated material points can be placed in a K-D tree (Samet 1990) for efficient searching. However, this tree must be rebuilt each time the mesh is updated (e.g. during iterative mesh fitting). As the density of the material points increase, $\epsilon_{DM}$ converges towards $\epsilon_{DP}$. 
4.3. MESH FITTING

Figure 4.11: Fitting error defined by the distance between each sampled material point and its closest data point.

Nearest Data Point

The third method is a nearest data point search. As in the nearest material point search, the mesh is discretised at material points. But now for each material point $m_j$, we find the closest data point $d_j$ (Fig. 4.10) so that

$$
\epsilon_{MD} = \sum_{i=1}^{D} \| \Omega(m_j) - d_j \|^2
$$

This is the most efficient method, where expensive non-linear searches are avoided, and a K-D tree needs only to be built once using the data points. In addition, outlier data points are ignored for closer points. The disadvantage is that distant but accurate data points can also be ignored. This can be mitigated by finding the $k$ nearest data points, and taking the average distance.

4.3.2 Smoothing Errors

Insufficient data points resulting in an under-constrained fit, or noisy data points can lead to mesh defects (Fig. 4.12). Sobolev smoothing is used to penalise high curvature within elements, and element boundary normal smoothing is used to encourage smoothness across Lagrange elements. Both methods decrease the occurrence of mesh self-intersection and creasing. When meshes are generated from a statistical model, the Mahalanobis distance of the mesh (Eq. 4.6) is used to smooth the mesh as a whole.
Figure 4.12: Mesh defects resulting from a lack of smoothing penalties. (a) shows a sharp crease at the element boundary. (b) shows a sharp crease within an element. Different penalty functions are used suppress these two types of defects.

**Sobolev Smoothing**

The second order weighted Sobolev norm (Terzopoulos, 1986) is calculated from the derivatives of the geometry of each element. For a curve with $n_e$ 1-D elements, it is defined as

$$
\epsilon_{sob}^{1D}(X) = \sum_{i=1}^{n_e} \int w_1 \sum_{j=1}^{3} \left( \frac{\partial \Omega_i^j}{\partial \xi} \right)^2 + w_2 \sum_{j=1}^{3} \left( \frac{\partial^2 \Omega_i^j}{\partial \xi^2} \right)^2 d\xi
$$

where $\Omega_i^j$ is the geometry in element $i$, in direction $j$.

For a mesh with $n_e$ 2-D elements, the Sobolev term is

$$
\epsilon_{sob}^{2D} = \sum_{i=1}^{n_e} \int \int w_{11} \sum_{j=1}^{3} \left( \frac{\partial \Omega_i^j}{\partial \xi_1} \right)^2 + w_{12} \sum_{j=1}^{3} \left( \frac{\partial \Omega_i^j}{\partial \xi_2} \right)^2 + w_{21} \sum_{j=1}^{3} \left( \frac{\partial^2 \Omega_i^j}{\partial \xi_1^2} \right)^2 + w_{22} \sum_{j=1}^{3} \left( \frac{\partial^2 \Omega_i^j}{\partial \xi_2^2} \right)^2 + w_{211} \sum_{j=1}^{3} \left( \frac{\partial^2 \Omega_i^j}{\partial \xi_1 \partial \xi_2} \right)^2 \partial \xi_1 \partial \xi_2
$$

The first order derivatives regularises the length of elements along $\xi$ directions, while the 2$^{nd}$ order derivative penalises against high curvature. The cross-derivative in the 2D case penalises against very large, or very small areas in 2-D elements (Bradley et al., 1997).
4.3. MESH FITTING

Figure 4.13: Tangent smoothing for 1D line elements (left). Normal smoothing for 2D surface elements (right)

Element Boundary Smoothing

Piece-wise parametric meshes with Lagrange basis functions are $C^0$ continuous, that is, they are only continuous in position, and not tangent direction or curvature across elements. However, for geometric accuracy, and visual aesthetics, it is desirable for these properties to be continuous. Thus penalty functions are applied against smoothness discontinuities at element boundaries.

For 1-D line elements, we penalise against differences in tangent directions at element junctions. Given a 1-D mesh with $n_t$ connections between line elements, the tangent smoothing penalty is:

$$
\epsilon_b^{1D} = \sum_{i=1}^{n_t} 1 - t_{i,1} \cdot t_{i,2}
$$

(4.19)

where $t_{i,1}$ and $t_{i,2}$ are the normalised tangent vectors at the joining ends of the two elements at connection $i$. This is illustrated in figure 4.13.

For 2-D elements, differences in element boundary normals is penalised. Given a mesh with $n_b$ shared element boundaries, the penalty is:

$$
\epsilon_b^{2D} = \sum_{i=1}^{n_b} \int 1 - N_{i,1} \cdot N_{i,2} d\xi_b
$$

(4.20)

where $N_{i,1}$ and $N_{i,2}$ are normalised mesh surface normals evaluated in the two adjacent elements along the shared boundary $i$, with element coordinates $\xi_b$.

Mahalanobis Distance

Given a set of mesh parameters $X'$ generated from a statistical model, such as a set of principal components (Sec. 4.1), the Mahalanobis distance (Eq. 4.6) quantifies the
similarity of $X'$ to the training set of the model. The higher the Mahalanobis distance, the more dissimilar the generated parameters are from training set parameters. Assuming that training set meshes are free of defects, generated parameters that cause mesh defects will produce a high Mahalanobis distance. See section 4.3.3 for details regarding parameter generation from principal components.

### 4.3.3 Fitting Methods

Having defined fitting errors, this section presents methods for adjusting mesh parameters to minimise those errors.

**Nodal Fit**

In a nodal fit, all mesh parameters are free to adjust independently to minimise fit error in a least-squares sense. Nodal fits are computationally expensive because of the high number of degrees of freedom, and require careful tuning of smoothing weights. However it produces the most accuracy fits.

**Free-Form Deformation**

Free-Form Deformation (FFD), or host-mesh fitting [Fernandez et al., 2004], is used to fit a mesh on a coarse scale. As shown in figure 4.14, the mesh (slave mesh) is embedded in a coarse host mesh with fewer degrees of freedom. Fitting adjusts the parameters of the host mesh, and as the host-mesh deforms, the slave mesh deforms according to the warping of the space internal to the host mesh.

The host mesh is typically composed of a small number of low-order tri-cubic elements that completely envelop the slave mesh. The slave mesh is embedded by calculating the host mesh material coordinates of the slave mesh nodes (slave nodes). As the host mesh deforms, slave node coordinates are updated by evaluating the host mesh at these material coordinates.

Since the host mesh has far fewer degrees of freedom than the slave mesh, the complexity of the fit is greatly reduced at the cost of fitting accuracy. FFD is used to bring a template mesh close to the data efficiently before nodal fitting fine tunes the fit.
4.3. MESH FITTING

Figure 4.14: Free-form deformation, or host-mesh fitting of a rectus femoris muscle. The mesh of the muscle is embedded as a slave mesh in a simple host mesh. Material points on the slave mesh are fitted to data points by deforming the host-mesh, which deforms the embedded slave mesh. Reproduced from Fernandez et al. (2004).

Principal Component Fit

Figure 4.15: Proximal femur mesh change along a principal component. Mesh geometry can be customised by adjusting the weighting along each component.

Given a set of $k$ principal components (PCs) trained from a set of meshes, mesh parameters $\mathbf{x}'$ can be generated from the weighted sum of the principal components $\mathbf{b}$, and the mean $\bar{\mathbf{x}}$. This is described in equation 4.5, which is repeated here:

$$
\mathbf{x}' = \bar{\mathbf{x}} + \sum_{i=1}^{k} \mathbf{b}_i a_i
$$

(4.21)
By adjusting the weights (Fig. 4.15), or PC scores \(a\), plus a rigid body translation \(t\) and rotation \(r\) (and scaling if training meshes are not normalised by size), a mesh can be customised to minimise error. In chapter 7, the first five PCs are sufficient to fit a femur mesh to segmented data, giving a total of only 11 degrees of freedom in the fit.

PC fitting has three main advantages over the previous two methods: one, far fewer degrees of freedom are involved, meaning more efficient fitting; two, it is robust against unrealistic mesh geometries or mesh defects, since within a certain range of weights, \(x'\) is representative of the meshes used to train the principal components. To constrain the weights to realistic ranges, the Mahalanobis distance of the reconstruction is used as a smoothing error (Sec. 4.3.2); and three, PC fitting distributes mesh nodes in a way that is correspondent across the population (assuming correspondence in the training set), so that iterative PC-fitting and training can be used to improve the correspondence of fitted meshes (see Sec. 6.3). These advantages rely on the assumption that the PCA training set is normally distributed so that meaningful variation is captured in the first few components. Chapters 6 and 7 show that this assumption is valid for normal femurs, and that only 5 components are needed to capture the majority of femur shape variation.
4.4 Active Shape Modelling

Active shape modelling (ASM) was introduced in (Cootes et al., 1995) as an application of statistical shape modelling to image segmentation. Despite its age, ASM is still widely used (Heimann et al., 2006; Gregory et al., 2007; Sun et al., 2010) due to its simplicity and efficiency. The femur CT-to-model pipeline uses ASM to provide an initial segmentation of the femoral surface (Chap. 7). Details and discussions of the original implementation can be found in (Cootes et al., 2001). Here, we present the ASM methodology adapted for piece-wise parametric meshes.

4.4.1 Training

ASM requires a statistical model of object shape, and models of image appearance normal to the object surface (local appearance models). Shape model training was discussed in chapter 2, so will not be repeated here. Local appearance models consist of statistical models of normalised image gradient profiles across the surface of the object. The models are trained on a set of images and their segmented surfaces. For the rest of this section, we will assume the surface to be described by a piecewise parametric mesh.

For each image, image intensity profiles are sampled normal to evenly spaced mesh material points, $q^{ASM} = \{q_i^{ASM} | i = 1, \ldots, Q^{ASM}\}$ (Fig. 4.16a). These profiles extend a short distance either side of the mesh along the mesh normal. The gradients of the profiles are computed, and normalised. Normalising the gradient of each profile eliminates variations in global intensity between samples from different images. As proposed in Cootes et al. (2001), the gradient of a profile with 2l sample along its length, denoted as a vector $g'$, is normalised to $g$ by:

$$g = \frac{g'}{\sum_{i=1}^{2l} |g'_i|}$$

where $g'$ is divided by the sum of the absolute values of its element. As an example, average profiles from the femoral head and shaft are shown in figure 4.16b.

At each material point $q^{ASM}$, PCA on $\{g_i | i = 1, \ldots, m\}$ from across the training set produces a mean profile, principal components $\{b_i | i = 1, \ldots, m - 1\}$ and eigenvalues $\{\lambda_i | i = 1, \ldots, m - 1\}$. $m$ is the size of the training set. The Mahalanobis
Figure 4.16: Illustration of image profile sampling for ASM training. Image profiles are sampled normal to material points on the mesh. (a) shows the average normalised gradient profiles at two material points: femoral head (above), and femoral shaft (below). Error bars show one standard deviation from the mean at each sample point. For the femoral head, the rise in value past the half-way point of the profile is due to the pelvis. The shaft shows a clean profile with only soft-tissue either side of the cortex.

The Mahalanobis distance is a measure of $g^*$’s similarity to profiles from the training set. The smaller $h$, the greater the similarity. During ASM segmentation, the profile models are used to identify probable matches for the femoral surface in the image.

4.4.2 Segmentation

The goal of ASM segmentation is to produce a mesh $\Omega$, with parameters $X$, that approximates the in-image object surface. The surface is predicted by the local appearance
Figure 4.17: Surface prediction along a sampled profile in ASM segmentation. A 1-D profile of length $2(l + l')$ is sampled from the image, portions of length $2l$ are used to calculate Mahalanobis distances against the local appearance model to quantify its likelihood of being from the object surface.

models at material points. $X$ is produced from a rigid-body translation $u = (u_x, u_y, u_z)$ and rotation $r = (r_x, r_y, r_z)$ of mesh parameters, generated by the shape model with principal component weights $a = (a_1, \ldots, a_{np})$. Thus $\Gamma = (u, r, a)$ is optimised to give $\Omega$ that matches the predicted surface. For convenience, $\Omega(\Gamma)$ will denote the mesh with parameters produced by $\Gamma$, and

$$p_i = \Omega(q_i^{ASM}, \Gamma)$$

is the $\mathbb{R}^3$ coordinate of $\Omega(\Gamma)$ evaluated at material point $q_i^{ASM}$ ($i = 1, \ldots, Q^{ASM}$) (these are the same as those with local appearance models).

Initial $\Gamma$ is provided by some initial estimate of the position and orientation of the object. Iterative segmentation follows, where surface prediction, and optimisation of $\Gamma$ are repeated until convergence.
Surface Prediction

In the surface prediction step, the image is searched normal to the mesh surface to find the best match for each local appearance model (Fig. 4.17). The image is sampled normal to the mesh at the ASM material positions $q^{ASM}$ where the profile models are trained. A sampled gradient profile $g^*$ is longer than the model profiles but of the same resolution. Given a length of $2l$ samples for model profiles, $g^*$ is of length $2(l + l')$. Sub-profiles $g^*_i = (g^*_{i}, ..., g^*_{i+2l})$ can be extracted from $g^*$, with $i$ ranging from 0 to $2l'$.

Calculating the Mahalanobis distance $h_i$ of each $g^*_i$ to the profile model (Eqn. 4.23) gives the vector $h = (h_1, ..., h_{2l'})$. Recall that the smaller the Mahalanobis distance, the more similar a profile is to the profile model, making $g^*_{\arg\min(h)}$ the best matching profile. By construction, the object surface should be at the centre of a profile, so the matched position of the surface along the profile is $d = \arg\min(h) + l$. This 1-D coordinate is converted into 3-D coordinates $d$ by

$$d = \rho \left[ d - (l + l') \right] n + p$$

where $\rho$ is the resolution of profile and $n$ is the surface normal at a point with coordinates $p$ on the mesh. Once $d$ has been calculated for each ASM material point, we obtain a cloud of data points $D = \{d_i \mid i = 1, \cdots, Q^{ASM}\}$ representing the predicted object surface.

Element-wise Error Correction

Variations in bone morphology, joint positioning, or image artefacts can cause errors in the surface prediction (Fig. 4.18a). However, one can assume that correctly predicted surface points will closely follow the mesh surface, which is constrained to realistic geometries by the shape model. Thus the match positions $d$ along sampled profiles should be similar for all point in a mesh element.

Over each element, profiles with $d$ greater than one standard deviations away from the median $\tilde{d}$ are flagged as possibly erroneous. For each of these profiles, all possible match positions (all local minima of $h$) are found, and the match closest to $\tilde{d}$ is selected as the corrected match position. The 3-D coordinates of these matches are recalculated based on the corrected $d$. 
Mesh Update

In the mesh-update step, the mesh is fitted to the predicted surface data cloud \( \mathbf{D} \) by adjusting shape model model weights and rigid transforms. The objective function is:

\[
\min_{\Gamma} \epsilon_{\text{ASM}} = \sum_{i=1}^{Q_{\text{ASM}}} \| \mathbf{d}_i - \Omega(q_{i,\text{ASM}}^*, \Gamma) \|_2^2 + h_{\text{mesh}}(\Gamma) 
\]

where \( h_{\text{mesh}} \) is the Mahalanobis distance of the current mesh, with shape model weights \( \mathbf{a} \) (which are a part of \( \Gamma \)). \( h_{\text{mesh}} \) penalises against unlikely mesh shapes being produced during the fit.

The next segmentation iteration is performed using the updated mesh geometry, and so on until any of the following conditions are satisfied:

1. A maximum number of iterations is reached;
2. When the optimum mesh parameters \( \Gamma \) have not changed between iterations;
3. When at least 80% of \( q_{i,\text{ASM}}^* \) have their match positions within the central 20% of their sampled profiles.

Criteria one terminates segmentations that do not converge, which occur when the mesh is too far from the object of interest. Criteria two is satisfied when the predicted mesh has converged. This does not guarantee that the mesh has converged to the correct shape, but since the mesh is not changing, no further improvements will
be made. The third criteria is satisfied when the mesh has converged to the correct shape, within tolerances based on suggestion in [Cootes et al. (2001)]. The tolerances are relatively loose because the statistical shape model is not to expected to closely match all objects of interest due to limitations of the training set and the Mahalanobis distance penalty. A subsequent mesh fit (Sec. 4.3) to segmented points $\mathbf{D}$ can fine tune the mesh geometry without the constraints of a shape model.
4.5 Cortical Thickness Mapping

Cortical Thickness Mapping (CTM) allows sub-pixel estimation of cortical bone thickness from CT images. The femur CT-to-model pipeline uses CTM to finely segment the femoral cortex following ASM segmentation (Ch. 7). As shown in figure 4.19, CTM involves modelling the image intensity profile across the cortex by a function $y_{\text{blur}}$. $y_{\text{blur}}$ is the piece-wise constant function $y$ convolved by two point spread functions simulating in-plane blurring $G_i$ and out-of-plane blurring $G_o$. The piece-wise constant function $y$

$$y(c) = y_0 + (y_1 - y_0) H \left( c + \frac{c_1 + c_0}{2} \right) + (y_2 - y_1) H \left( c + \frac{c_1 - c_0}{2} \right)$$

(4.27)

is composed of three sections representing the intensity values of trabecular bone $y_0$, cortical bone $y_1$, and external tissue $y_2$ with the cortical surfaces at $c_0$, $c_1$. $H$ is the unit step function.

The in-plane blurring function $G_i$ is a normalised Gaussian

$$G_i(c) = \frac{1}{\sigma \sqrt{\pi}} e^{-\frac{c^2}{2\sigma^2}}$$

(4.28)
controlled by a single parameter $\sigma$.

$G^o$ is a rectangular function

$$G^o(c) = \frac{1}{2r} [H(c+w) - H(c-r)]$$

(4.29)

with its half-width $w = \frac{1}{2} s \tan \alpha$ governed by the angle $\alpha$ between the normal and the image slice plane, and the slice spacing $s$. Modern CT systems allows images to be re-sliced with arbitrary orientations and spacing, such that an ellipsoidal 3-D point spread function may be more appropriate. Discussions with the authors of Treece et al. (2010) revealed that CTM was was insensitive to whether a Gaussian and rectangular or ellipsoidal filter was used. For this work, the original published and validated method using a rectangular $G^o$ is implemented.

In summary, $y_{\text{blur}}$ is given by

$$y_{\text{blur}}(c) = y_0 + \int \int \frac{y_1 - y_0}{2w\sigma\sqrt{\pi}} \left[ \exp \left( \frac{c + w - c_0}{\sigma} \right)^2 - \exp \left( \frac{c - w - c_0}{\sigma} \right)^2 \right] + \frac{y_2 - y_1}{2w\sigma\sqrt{\pi}} \left[ \exp \left( \frac{c + w - c_1}{\sigma} \right)^2 - \exp \left( \frac{c - w - c_1}{\sigma} \right)^2 \right] dc dc$$

(4.30)

### 4.5.1 Model Fitting

The model $y_{\text{blur}}$ is fitted to a sampled profile $\hat{y}$ of length $l$, by optimising $\Delta = (c_0, c_1, y_0, y_2, \sigma)$. The objective function is

$$\min_{\Delta} \epsilon_{\text{CTM}} = \int w^{\text{CTM}}(c) [\hat{y}(c) - y_{\text{blur}}(c, \Delta)]^2 dc$$

(4.31)

The weighting function $w^{\text{CTM}}(c)$ serves two purposes. One, to focus the fit on the cortex, and two, to suppress noise from non-cortical features such as trabecular and neighbouring bones. The choice of weighting functions is detailed below. Cortical density is assumed to be constant over the bone, so $y_1$ is fixed to a value estimated either from a thick portion of the cortex, with CT values unaffected by partial-volume effects; or from the distribution of estimated thickness values (Treece et al. 2012).

### 4.5.2 Weighting Function

Two weighting functions are used depending to the environment around the cortex. Where there are no features around the cortex (e.g. the femoral shaft), $w^{\text{CPM}}(c)$ is the
4.5. CORTICAL THICKNESS MAPPING

Figure 4.20: Fitted CTM profiles with weighting functions in black. In the shaft, weights are calculated from the gradient of the sampled profile \( a \). In the proximal and distal femur regions, weights are given by equation 4.32 \( b \).

Gaussian gradient magnitude of the sampled profile (Fig. 4.20a). This results in the rising and falling edges of the cortical boundaries being highly weighted.

Where non-cortical features are present (e.g. joints), \( w^\text{CPM}(c) \) is a double-peaked function

\[
w^\text{CPM}(c) = \frac{c + c_0}{500 (c - c_0)^2 + (c + c_0)} + \frac{c + c_1}{500 (c - c_1)^2 + (c + c_1)} \quad (4.32)
\]

centred about the initial estimate of the cortex at \( c_0 \) and \( c_1 \) (Fig. 4.20b). Features around the cortex are suppressed, and the rising and falling edges of the cortical boundaries are again highlighted.
4.5.3 Inner and Outer Cortical Surface Segmentation

Cortical thickness is simply calculated as \( t = c_1 - c_0 \). In addition, \( c_0 \) and \( c_1 \) are converted into the 3-D coordinates of points \( d^{\text{in}} \) and \( d^{\text{out}} \), on the inner and outer cortical surfaces, respectively, according to:

\[
\begin{align*}
d^{\text{in}} &= c_0 n - p \\
d^{\text{out}} &= c_1 n + p
\end{align*}
\]

where \( n \) is the surface normal vector at a point with 3D coordinates \( p = \Omega^{\text{ASM}}(q^{\text{CTM}}) \).
4.6 Radial Basis Functions

Radial basis functions (RBFs) provide a meshless way of smoothly approximating or interpolating spatially varying quantities, or fields. Following femoral cortex segmentation, the femur CT-to-model pipeline uses RBFs to register cancellous bone mineral density distribution (Cha. 8). Unlike mesh-based approximations, such as B-splines, or finite-element meshes, RBF fields require no connectivity information, and RBF knots can be placed arbitrarily. RBFs have been applied to a number of computational problems, including sparse data interpolation (Pouderoux et al., 2004; Oubel et al., 2010), neural-networks (Orr et al., 2000; Leonardis and Bischof, 1998), computer graphics and imaging (Magoulès et al., 2007; Vaněček, 2010; Süßmuth et al., 2010; Carr et al., 2001), numerical modelling (Bah et al., 2009; Jakobsson and Amoignon, 2007; Larsson and Fornberg, 2003), and image registration (Siddiqui et al., 2009; Fornefett et al., 1999; Wachowiak et al., 2004; Matsopoulos et al., 2005).

A radial basis function $\phi$ is a function who’s value at $d$ is only dependent on the distance of $d$ from the knot $c$ of the function, such that

$$\phi(d) = \phi(||d - c||)$$

(4.33)

The RBF field value $u$ at $d$ is the weighted sum of multiple RBFs with knots $c_1..Q$:

$$u(d) = \sum_{i=1}^{Q} a_i \phi_i(r_i)$$

(4.34)

where $r_i = ||x - c_i||$, and $a_1..Q$ are the weights associated with each of the knots.

To approximate a field with known values $u = (u_1, ..., u_D)$ at $d_1, ..., D$, $a = (a_1, ..., a_Q)$ is fitted according to the linear system

$$Ha = u$$

(4.35)

where

$$H = \begin{pmatrix} \phi_1(r_{1,1}) & \cdots & \phi_N(r_{1,Q}) \\ \vdots & \ddots & \vdots \\ \phi_1(r_{D,1}) & \cdots & \phi_N(r_{D,Q}) \end{pmatrix}$$

(4.36)
Figure 4.21: Thin-plate spline RBF approximation (centre) of a 2-D function (left). Fifty randomly located knots (green circles) were fitted to 500 randomly sampled data points (clear circles).

Table 4.1: Common radial basis functions. + denotes functions only defined between $0 \leq r \leq 1$, and is 0 elsewhere.

<table>
<thead>
<tr>
<th>Type</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>biharmonic</td>
<td>$\phi(r) = r^2$</td>
</tr>
<tr>
<td>thin-plate spline</td>
<td>$\phi(r) = r^2 \log r$</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$\phi(r) = \exp\left(-\frac{r^2}{\sigma^2}\right)$</td>
</tr>
<tr>
<td>Wendland’s</td>
<td>$\phi(r) = (1 - r)^2$</td>
</tr>
</tbody>
</table>

in which $r_{i,j} = \|d_i - c_j\|$. Figure 4.21 illustrates fitting 50 randomly distributed knots to 500 randomly sampled points of a 2-D function. Maximum error is 0.25, with an RMS error of 0.06.

4.6.1 Common Functions

Table 4.1 list common functions $\phi$. Functions are classified as either of local support, or global support. Local (or compact) support functions are zero beyond a certain values of $r$. An example is the Wendland function [Wendland 1995]. Local support functions have been used for image registration, for their ability to apply local deformations at arbitrary locations [Siddiqui et al. 2009; Fornellet et al. 1999; Wachowiak].
Global support functions are non-zeros as \( r \) tends to infinity, e.g. the popular Gaussian function. According to Buhmann (2003), global support functions generally give better approximation accuracy, at the cost of numeric-stability when directly solving equation 4.35.

4.6.2 Gaussian RBF Width

Gaussian RBFs are used for image registration in chapter 8. The Gaussian width \( \sigma \) is an important parameter that governs the overlap between each RBF knot, which affects approximation accuracy. A constant \( \sigma \) for all knots can lead to sub-optimal overlap when knot distribution is non-uniform. We follow the methods of Moody and Darken (1989) and Benoudjit et al. (2002) in determining the width for each knot based on its distance to its \( k \) nearest knots, and scaling factor \( s \):

\[
\sigma_i = s \frac{1}{k} \left( \sum_{j=1}^{k} \| c_i - c_j \|_2^2 \right)^{\frac{1}{2}}
\]  

(4.37)

4.6.3 Fitting Methods

Regardless of the choice of basis function, fitting RBF fields by directly solving equation 4.35 is known to be an ill-conditioned problem (Fornberg et al., 2011). In addition, it possesses \( O(Q^3) \) complexity, making it inefficient, and prohibitive when \( Q \) is greater than a few hundred.

Local support functions produce sparse \( H \) matrices that can take advantage of efficient solving methods (Saad, 2003), at the cost of approximation quality (Torres and Barba, 2009).

Global support functions tend to give more accurate approximations (Buhmann, 2003), but produce a fully populated \( H \) that can be ill-conditioned. However, methods such as QR decomposition (Fornberg et al., 2011), orthogonal least-squares (Chen et al., 1991), and far-field approximation (Torres and Barba, 2009), can be used to overcome instabilities in solving the fitting problem.
4.6.4 Evaluation Methods

Evaluating a RBF field at a point requires $Q$ operations to evaluate the influence of $Q$ RBFs, so the evaluation of $M$ points is $O(Q \times M)$. However, fast evaluation methods exist that can perform field evaluations in $O(N)$ operations. For Gaussian basis functions, there is the Fast Gauss Transform method (Greengard and Strain, 1991). For basis functions with wider support, such as multiquadrics, there is the Fast Multipole method (Greengard and Rokhlin, 1987), which has $O(Q^2)$ complexity.

4.7 Summary

The various computational methods described in this chapter are used in the femur CT-to-model pipeline for femur shape description (Lagrange meshes), shape model training (mesh fitting, PCA), automatic segmentation (ASM, CTM), and automatic image registration (RBF). As the pipeline usage of these methods are described and discussed in following chapters, relevant sections here will be frequently referenced for readers looking for details.
Chapter 5

VIFM CT Images

The CT-to-model pipeline for femur morphology was developed with 262 post-mortem quantitative computed tomography (QCT) images from the Victorian Institute of Forensic Medicine (VIFM). The images came from a total of 327 images collected for the study by Hislop-Jambrich (2010). Each image was complemented with gender, age-at-death, height, and weight information.

Of the 262 images, 41 were used for the training phase of the pipeline, and 17 were used to validate pipeline accuracy. The complete set of images was used to test pipeline robustness.

This chapter will present details about the VIFM images and their demographic information. A brief introduction to the VIFM is first presented in the next section, followed by ethical approval details. Imaging protocols and characteristics are presented in section 5.3 and demographic information is presented in section 5.4. Finally, section 5.5 briefly discusses how the dataset was used to develop the CT-to-model pipeline.

Two appendices are attached to the end of this chapter. Appendix 5.A documents the laser-scanning of 10 excised proximal femurs from the Melbourne Femur Collection (MFC). Appendix 5.B outlines CT theory, scanner technology, and the technical details of QCT for readers unfamiliar with these topics.
5.1 Background

The Victorian Institute of Forensic Medicine (VIFM) is the statutory provider of forensic pathology, clinical forensic medicine, and tissue banking in the state of Victoria, Australia (The Victorian Institute of Forensic Medicine, 2011). As a part of its autopsy service in its mortuary, the VIFM performs about 4700 post-mortem full-body CT scans per year (The Victorian Institute of Forensic Medicine, 2009). The images are used to augment traditional autopsy by radiologists trained in forensic radiology.

Between December 2007 and June 2008, the femur portion of the images from 327 individuals were selected for a study into the femur morphology of a modern Australian population (Hislop-Jambrich, 2010). All individuals were from the state of Victoria, and most passed away due to unexpected causes. Access to this dataset was given by the VIFM and The University of Melbourne, to develop an automated femur-morphology extraction pipeline. The intention was to use the pipeline to collect femur morphology from the thousands of images in the VIFM database, or others like it. This would allow accurate, and representative inferences regarding the variations in femur morphology in a modern population.

5.2 Ethical Approval

CT data and femur bone samples were collected by (Hislop-Jambrich, 2010) under existing ethics applications. The current author was added to the university applications after the data had already been collected. The ethics applications allowed the use of the anonymised data for the development of the CT-to-model pipeline, and for future research.

The VIFM dataset was collected under ethics application EC9/2007 (prospective), and EC10/2007 (retrospective). The MFC proximal femur portions were collected under ethics applications EC26/2000 and HREC 980139.
5.2. ETHICAL APPROVAL

Figure 5.1: Typical VIFM images. Figure (a) is a maximum intensity projection in the coronal plane, showing the extent of the image in the superior-inferior direction. Figure (b) is a transverse slice showing the body lying above the QCT phantom and the scanner gantry, as labelled.
5.3 Image Description

As defined by the ethics applications, only the region containing the femur was collected from each full-body image. This region contains full axial slices from just superior of the pelvis, to just inferior of the knee (Fig. 5.1a). Except in rare cases, the visible anatomic features are the pelvis, both femurs, both patella, and the superior end of the tibia and fibula, and surrounding soft tissue. In addition, the CT scanner gantry and the QCT phantom are also visible (Fig. 5.1b). See appendix 5.B.4 for details regarding the phantom. In many cases, portions of the hands, and various objects such as catheters, or prostheses, are also visible. The effect of these are discussed in section 5.3.2.
5.3. IMAGE DESCRIPTION

Figure 5.3: The VIFM CT scanning room. The ceiling-mounted body-lift is on the left, while the Toshiba Aquilion CT scanner is visible on the right. Reproduced from [Hislop-Jambrich (2010)]

5.3.1 Imaging Setup

The VIFM CT scanner was a commercially available Toshiba Aquillion 16 MDCT\textsuperscript{1}. The Mindways QCT\textsuperscript{2} system was used in conjunction for bone mineral density measurements. During the imaging procedure, a body was manually placed on the scanning table by a technician, with the help of a body-lifter (Fig. 5.3). The positioning of the body on the table was not strictly controlled, so a wide range of limb positions and orientations were imaged.

In the mortuary environment, X-ray dosage was not an issue, so higher-than-normal X-ray energies were used for more robustness against variable conditions and operator training (for more details regarding CT theory, refer to section 5.B). Machine settings

\textsuperscript{1}Toshiba Medical Systems Corporation, Tokyo, Japan. http://www.toshiba-medical.eu/en/Our-Product-Range/CT/Systems/Aquilion-16/

\textsuperscript{2}Mindways Software, Inc. Austin, TX, USA. http://qct.com/
Table 5.1: VIFM CT scanner parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>scan type</td>
<td>helical</td>
</tr>
<tr>
<td>pitch factor</td>
<td>1.0</td>
</tr>
<tr>
<td>slice thickness</td>
<td>2.0 mm</td>
</tr>
<tr>
<td>slice spacing</td>
<td>1.6 mm</td>
</tr>
<tr>
<td>slice overlap</td>
<td>0.4 mm</td>
</tr>
<tr>
<td>pixel spacing</td>
<td>0.976 mm x 0.976 mm average</td>
</tr>
<tr>
<td>typical kVp</td>
<td>120</td>
</tr>
<tr>
<td>typical tube current</td>
<td>180 mA (adjustable)</td>
</tr>
<tr>
<td>typical FoV</td>
<td>500 mm</td>
</tr>
<tr>
<td>image size</td>
<td>512 x 512</td>
</tr>
<tr>
<td>typical resolution</td>
<td>1.025 pixels per mm (averaged)</td>
</tr>
<tr>
<td>focal spot size</td>
<td>1.6 mm x 1.4 mm</td>
</tr>
<tr>
<td>kernel</td>
<td>FC01 body</td>
</tr>
<tr>
<td>patient position</td>
<td>supine feet first</td>
</tr>
<tr>
<td>typical number of slices</td>
<td>400</td>
</tr>
</tbody>
</table>

used for scanning are listed in table 5.1.

A reconstruction algorithm was selected to provide better low contrast resolution at the cost of spatial resolution ([Hislop-Jambrich, 2010](#)). This was because for post-mortem examinations, it was important to be able to differentiate between different soft tissues. For privacy reasons, image meta-data was limited to age-at-death, gender, height, and weight, as per the ethics application.

### 5.3.2 Image Artefacts

The VIFM mortuary presented a relatively uncontrolled imaging environment compared to clinical settings. As a result, some images contained severe image artefacts that prevent reliable image analysis. The cause of the artefacts were mainly to do with body positioning and the presences of metallic objects in the field of view, as illustrated in figure [5.4](#).

Bodies were commonly scanned while inside a plastic body bag. This was to protect the scanner, and to preserve the state of the body. For subsequent coroner examinations, any objects on or inside the body at the time of death were not removed for scanning. As a result, medical equipment such as catheters were present in many image. More problematic however, were metallic objects such as staples, pins, screws, and
metallic implants in the hip and knee. These objects have extremely high attenuation coefficients, which resulted in severe beam-hardening and bright streak-like artefacts. Another artefact was caused by limbs lying on the edge of the CT field of view. This cropping artefact distorted CT values across axial slices (Fig. 5.5).

The artefacts described above prevented accurate analysis of the images. Streak-like artefacts obscured anatomical features, making geometric measurements unreliable. Beam hardening and cropping artefacts changed the relationship between CT number and tissue density. This compromises the accuracy of BMD measurements via QCT. The image selection criteria for Hislop-Jambrich (2010) (see next section) aimed to
avoid images with these artefacts. However, in order to collect a sufficient number of images, images with limited artefacts were included when it was possible to mitigate their effect.

### 5.4 Data Demographics

The VIFM dataset consisted of 327 Australian individuals from the state of Victoria, who died unexpectedly. 169 were male, and 158 were female. The images were selected based on having a normal femur. Images were excluded in cases where the following was true:

1. Femur showed trauma or pathology unrelated to normal ageing;
2. Positioning of the body caused gross distortions or artefacts;
3. Homicide cases, or had other ethical issues (e.g. well-known people).

An attempt was made to collect 20 images per 10-year age-group, from under 20 to above 80 years of age. However, there was difficulty in finding the required number
Table 5.2: Average age of death, weight and height for the VIFM dataset. Standard deviation is given in brackets.

<table>
<thead>
<tr>
<th></th>
<th>combined</th>
<th>male</th>
<th>female</th>
</tr>
</thead>
<tbody>
<tr>
<td>age of death (years)</td>
<td>57.8 (21.5)</td>
<td>55.8 (21.5)</td>
<td>59.9 (21.2)</td>
</tr>
<tr>
<td>weight (kg)</td>
<td>71.7 (17.2)</td>
<td>76.5 (16.5)</td>
<td>66.5 (16.3)</td>
</tr>
<tr>
<td>height (cm)</td>
<td>166.1 (9.7)</td>
<td>172.1 (7.6)</td>
<td>159.6 (7.2)</td>
</tr>
</tbody>
</table>

of image for some age groups, particularly the younger groups.

Table 5.2 summarises the averages and standard deviations of age-at-death, body weight, and body height of the dataset as a whole and by gender. Figures 5.6, 5.7, and 5.8 show the distribution of age of death, weight and height, respectively.

As can be seen from figures 5.7 and 5.8, weight and height distributions were normal, with males about 15% heavier and 8% taller than females on average. While a positive relationship between height and weight might be assumed, Hislop-Jambrich (2010) showed that there was no evidence of such a relationship. Mean male age was slightly lower than that of females, due to a cluster of female subjects in the 70 to 90 years old range.

No other subject information was recorded. The ethnicity of each individual was not recorded for privacy reasons, but can reasonably be assumed to be predominantly white. A large number of subjects died unexpectedly from suicides and atypical accidents, which may bias the population towards a particular social group, particularly for younger subjects.

5.5 Use of the VIFM Dataset

The scope of the femur CT-to-model pipeline is limited to normal femurs and femur images. Images containing severe artefacts, or deformed, damaged or diseased femurs were removed from the dataset, leaving 262 images. Out of these, 41 images were used as a training set for the training stage of the pipeline. In addition, manual segmentations of the 17 MFC images were used to validate the accuracy of the pipeline’s segmentation step.
Figure 5.6: Age distribution of VIFM data by gender.

Figure 5.7: Weight distribution of VIFM data, by gender.
5.5. USE OF THE VIFM DATASET

5.5.1 Excluded Images

In total, 327 images were collected. However, for the purpose of this thesis, scans that contained abnormalities that would severely affect segmentation, or BMD analysis, were excluded. These abnormalities were

- Metallic objects: prosthesis, surgical equipment, jewellery (37 images);
- Abnormal morphologies: down syndrome, fractures, tumours (4 images);
- Severe osteoporosis and osteoarthritis: very low bone density, spurs, porous cortex (8 images);
- Clipping artefacts: image artefact from body extending beyond imaging field of view (6 images);
- Truncated stacks: image stack did not encompass whole femur (10 images).

Figure 5.9 provides examples of these. In total 65 scans were excluded, leaving 262 scans for pipeline processing.

5.5.2 Training and Validation Set

As outlined in the introductory chapter (Cha. 1), the pipeline contains a training stage in which an initial femur statistical shape model is trained. A training set of 41 images
Figure 5.9: Abnormalities that excluded images from pipeline processing: (a) prosthesis and metallic objects, (b) fractures and deformities, (c) severe osteoporosis and osteoarthritis, (d) clipping artefact, (e) truncated image stack.
were randomly selected for this purpose. Also, a set of ground-truth morphologies were required to validate the accuracy of the pipeline. Seventeen images were set aside for this as the validation set.

Figure 5.10: Segmenting the training and validation set images. First, non-femur objects were deleted \([b]\) then the femur surface was segmented using edge-detection and manual editing \([c]\) to produce a femur surface point cloud \([d]\).

The outer cortical surface of the left femur in each training and validation set image was manually segmented, the process of which is illustrated in figure 5.10 and detailed below:

1. The left femur in each image was manually cropped from the rest of the image, and adjacent bones were manually erased in ImageJ\(^3\);
2. Edge-preserving smoothing (e.g. median filtering) was applied to the image using the Insight Tool-Kit \([\text{Yoo et al.}, 2002]\) (ITK) Curvature Anisotropic Diffusion Image Filter. The filter was run for five iterations, with timestep = 0.05 and conductance = 1;
3. Edge detection was applied using the ITK Canny Edge Detection Image Filter, with a lower image intensity threshold of 20 to disregard edges in soft tissue. This step produced binary images of the cortical surfaces as well as other edge-like features inside the femur;
4. Gaps in the boundary of the femur was manually filled using ImageJ;
5. Region growing using the ITK Fast Marching Image Filter produced a binary im-

\(^3\) http://rsbweb.nih.gov/ij/
age that differentiated the whole femur volume from the background. Seed points for the filter were placed around the edges of the image, outside of the femoral surface. The filter was run for 800 time steps during which region growing propagated from the seed points. Due to the closed boundary of the femur, the region stopped at the femoral surface so that the interior and exterior of the femur were labelled as different regions;

6. The 3-D surface of the femur region was triangulated using the marching-cubes algorithm implemented in the Visual Tool-Kit\(^4\) (VTK) Contour Extractor function.

7. The triangle mesh was smoothed by the VTK Smooth Poly Data Filter to remove step-like voxel artefacts;

8. The triangle mesh was cleaned by the VTK Clean Poly Data function to merge very close vertices and collapse very short edges.

A degree of subjectivity existed in the manual segmentation process above, both from the manual corrections and choice of filter parameters. Subjectivity in the manual corrections was unavoidable as in any manual segmentation, but it was mitigated by the fact that all correction were made by the same operator. Filter parameters were chosen based on trial and error experimentation on a small subset of images, then applied to all images. Any systematic bias was likely a uniform enlargement of the femoral surface by about 1 voxel due to the smoothing operations. This bias was consistent across the training and validation image sets, so that errors calculated against the validation femurs should cancel out the bias. The uniformity and small magnitude of this bias was also unlikely to have a significant impact on the accuracy of shape variations exhibited by the training set femurs.

The manually segmented point clouds of the training set femurs provided the geometries for training the initial femur statistical shape model, which will be the topic of the next chapter. Manual segmentations of a further 17 femurs provided a set of ground-truth data for validating the accuracy of the pipeline’s automatic segmentation and cortex meshing step. This, and the use of other ground-truth data is discussed in chapter 7.

\(^4\)http://www.vtk.org/
Appendix 5
Seventeen images were complemented by excised left proximal femur portions. The proximal femurs are a part of the Melbourne Femur Collection (MFC). The MFC is one of the largest and best documented femur-tissue collections in the world. This 17 femur MFC set was composed of 11 males and 6 females, between the ages of 20 and 84 (mean age of 58.88 years). Mean height was 169.47 cm, and mean weight was 83.76 kg.

The surfaces of 10 MFC proximal femurs were digitised using a Konica Minolta\textsuperscript{5} laser scanner, with the goal of providing highly accurate ground-truth geometries for segmentation validation. However, to preserve bone quality and cartilage for future studies, the bones could not be thoroughly defleshed through chemical means. As a

\textsuperscript{5}Konica Minolta Holdings, Inc.. Tokyo, Japan.
result, the remaining connective tissue occluded the actually bone surface (Fig. 5.11). The occlusion was most severe in the femoral head (due to the presence of cartilage), and greater and lesser trochanters (major sites of muscle insertion). Unfortunately, the resulting digitised surfaces were deemed too far from the true bone surface for use as ground-truth data. However, they may be used in future work requiring femur models with cartilage geometry.

5.B X-ray Computed Tomography

X-ray computed tomography allows the femur to be imaged in-vivo. The combination of X-ray imaging with computer aided tomography gives rise to X-ray computed tomography (X-ray CT, or simply, CT). When X-ray is projected through a body to a detector, different tissues attenuate the radiation to different degrees. This allows the tissues to be differentiated by the intensity of the detected signal. Projections taken at many angles around the body can be reconstructed into a 2-D image, or slice, in the axial plane, using tomographic techniques. The intensity of each reconstructed pixel, measured in Hounsfield Units (HU), or CT number, represents the density of the tissue. A volumetric image can be constructed from a series of slice images.

Compared to the other major 3-D medical imaging modality, Magnetic Resonance Imaging (MRI), CT produces images of similar spatial resolution, but much higher temporal resolution, and is cheaper to maintain \cite{Pan2009}. In 2007, an estimated 72 million CT scans were performed in the United States \cite{Berrington2009}. The mineral composition of bone means that it give exceptional contrast in X-ray images. Therefore, bone density is routinely estimated from CT images.

In the rest of this section, the theory of X-ray CT will be explained in more detail. This provides technical background for the imaging characteristics of the VIFM dataset.

5.B.1 X-ray Imaging Theory

X-ray imaging relies on the attenuation of X-ray by tissue. Given a tissue between a X-ray source and a detector, X-ray passing through a tissue is attenuated by two main phenomena: Compton scattering, and the photoelectric effect, which are shown
in figure 5.12. In Compton scattering, a high energy photon (e.g. X-ray, or gamma ray) interacts with an atomic electron resulting in the photon being deflected, and losing some of its energy to the electron (Fig. 5.12a). In a photoelectric effect interaction, the energy of the photon is completely absorbed by the electron, which is ejected from the atom (Fig. 5.12b). In both cases, whether due to scattering or absorption, the number of photons reaching the detector is reduced, or attenuated. The energy of the photons are also reduced by Compton scattering.

The degree of attenuation of a tissue is summarised by its attenuation coefficient $\mu_{\text{tissue}}$. In general, the denser the tissue, the higher its coefficient. In X-ray CT, scanning parameters are calibrated so that water has a CT number of 0, and air of -1000, such that the CT value of the tissue in HU is

$$CT_{\text{tissue}} = \frac{\mu_{\text{tissue}} - \mu_{\text{water}}}{\mu_{\text{water}}} \times 1000$$  \hspace{1cm} (5.1)

How the CT number is used to quantify bone mineral density will be explained in section 5.B.4.

5.B.2 Tomography

Tomography is the process by which projections through an object from multiple angles is reconstructed into a planar or volumetric image. The mathematical method for tomographic reconstruction was pioneered by Johann Radon in 1917 (Radon [2005]). On a 2-D image, the Radon transform is the integral of any line through the image.
In conventional X-ray CT, these lines are obtained by projecting a fan-shaped beam through the target. Applying the Radon transform on multiple lines at many angles gives a sinogram (Fig. 5.13). Through the inverse Radon transform, the image can be reconstructed from its sinogram. However, computing the inverse transform following the original method involves solving an ill-conditioned linear system of equations.

The filtered-back projection algorithm offers a stable, and computationally efficient approximation of the inverse Radon transform. The algorithm is composed of two steps. First, each 1-D section is ramp-filtered, which can be efficiently performed by digital signal processing. Second, each filtered section is back-projected onto the image under reconstruction, along the direction it was taken. The intensity of each pixel is simply the sum of each back-projection. Variants of the filtered-back projection have superseded the inverse Radon transform, and is widely used today for computed tomography. Modern CT scanner use a cone-shaped beam instead of a fan-shaped beam, through the incorporation of multiple detector rows. This means that each projection is a 2D section through a 3-D volume instead of a 1-D section through a 2-D slice. Cone beam projection allow a 3-D volume to be imaged in a shorter time. The 3-D
variant of filtered-back projection is generally known as the Feldkamp algorithm \cite{Xiao2003, Sakamoto2005}. Proprietary variants of the Feldkamp algorithm are used in all modern scanners \cite{Hislop-Jambrich2010}.

More recent methods iteratively reconstruct the image based on a model of the expected image. These methods are more robust to noise, and need less projections for reconstruction, resulting in a lower X-ray dose \cite{Hislop-Jambrich2010}. Details about these tomography methods are beyond the scope of this work.

5.B.3 Projection Acquisition

As mentioned above, CT imaging require projections be taken at a number of angles around the target. Generally speaking, the more angles, the more accurate the reconstruction \footnote{exception being when very dense materials such as metals will still cause artefacts}. With this in mind, CT scanners are designed with the X-ray source and detectors at opposite ends of a ring, which is centred about the long axis of the target (Fig. 5.14). A large number of projections can be quickly taken as the source-detector ring spins. The projections lie on the same plane, and can be reconstructed into a 2-D or 3-D image (depending on the source beam shape). Then the target can be shifted incrementally along its long axis (z-axis), for another round of projections to reconstruct the next slice or volume. While conceptually simple, this approach is relatively slow. Modern CT scanners employ three methods to improve acquisition time for volume images.

The first method is the use of cone beams. As mentioned in the previous section, cone-beams produce a 2-D projections, which lead to the reconstruction of a volume, instead of just a single slice. This reduces the number of revolutions required to image a volume.

The second method is helical acquisition. Instead of a series of discrete circular paths around the target, the source and detector follow a helical, or spiral path around the target. In practice, this involves moving the target continuously through the source-detector ring as projections are taken, which reduced the time needed to cover the entire length of the target. However, this means that no set of projections discretely correspond to a particular slice. Instead, during reconstruction, projections
are interpolated along the axial direction as required. Indeed, the 'slice' can be at any location along the target’s length.

The third method is to use multiple detectors, giving rise to Multiple Detector CTs, or MDCTs. The source-detector ring carries an array of detectors, arranged along the axial direction. Combined with helical acquisition, a longer axial distance can be covered per revolution, and fewer revolutions are needed to image a volume.

The net effect of these three methods is that the target can be moved through the scanner at a faster rate, reducing acquisition time. It also means that there is a higher temporal resolution for imaging moving structures, such as the heart. In addition, the target is imaged as a continuous volume, rather than a discrete set of slices, so that the raw data can be resliced and resampled at will for different purposes during reconstruction.
5.B.4 Quantitative CT

Quantitative CT, or QCT, is an extension of X-ray CT that allows apparent Bone Mineral Density (BMD) to be calculated from CT number. There is a linear relationship between mineral density, and CT number. For a given CT image, this relationship can be calibrated by analysing the CT number of objects (phantoms) in the image with known mineral density. After the relationship is derived, it can be used to convert the CT number of any object (or tissue) in the image into an equivalent density value. QCT is primarily used to calculate apparent bone mineral density. The gold standard for measuring BMD is bone ash weight, which is the weight of mineral components after cremation. Compared to ash weight, QCT-derived apparent BMD has an accuracy of ± 10% \(^\text{(Hislop-Jambrich, 2010)}\).

The term *apparent* is used because the calculated density is not necessarily the true BMD due to the partial volume effect. The smallest spatial scale at which density is calculated is that of a voxel. At clinical CT resolutions, besides solid bone, a voxel volume can contain pores, marrow, or external tissue. The varying CT numbers of all these tissues are “lumped” into a single number. This is known as the partial volume effect. The calculated density is some average of these tissues, instead of just bone. Therefore, bone BMD estimates by QCT is more reliable in dense bone regions such as the cortex, than porous cancellous regions.

<table>
<thead>
<tr>
<th>Table 5.3: Mindways QCT phantom densities.</th>
</tr>
</thead>
<tbody>
<tr>
<td>phantom</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

The VIFM dataset used the Mindways QCT system\(^7\). The system assumed that the CT has been calibrated so that water is 0 HU, and air is -1000 HU. The calibration phantom consisted of five solid monomer rods (Fig. 5.15), each with a different attenuation coefficient equivalent to a known mixture of water and dipotassium phosphate.

\(^7\)Mindways Software, Inc. Austin, TX, USA. http://qct.com/
Figure 5.15: The Mindways QCT phantom, boxed in red.

\( \text{(K}_2\text{HPO}_4\text{)} \) (Tab. 5.3). \( \text{K}_2\text{HPO}_4 \) was used as an analogue for the mineral component of bone. The relationship between the CT number \( \mu \), the equivalent aqueous \( \text{K}_2\text{HPO}_4 \) density (KED) \( \rho_{\text{aq K}_2\text{HPO}_4} \), and the equivalent water density \( \rho_{\text{water}} \), of each rod was

\[
\mu = \alpha_{\text{ref}} \rho_{\text{aq K}_2\text{HPO}_4} + \rho_{\text{water}} + \beta_{\text{ref}}
\]  

(5.2)

Calibration involved sampling the CT number of each of the rods, then fitting the constants \( \alpha_{\text{ref}} \) and \( \beta_{\text{ref}} \) using linear regression. By adjusting \( \alpha_{\text{ref}} \) and \( \beta_{\text{ref}} \) for the removal of water

\[
\alpha_{\text{CT}} = \alpha_{\text{ref}} - 0.2174
\]

\[
\beta_{\text{CT}} = \beta_{\text{ref}} - 999.6
\]

the formula

\[
\rho_{\text{K}_2\text{HPO}_4} = \frac{\mu - \beta_{\text{CT}}}{\alpha_{\text{CT}}}
\]

(5.3)

was used to infer the solid KED of imaged tissue from its CT number. Note that KED is not a physical density, it is the density of \( \text{K}_2\text{HPO}_4 \) dissolved in water that would produce the observed CT number.
Chapter 6

Femur Shape Model Training

The first stage of the femur CT-to-model pipeline was training a femur statistical shape model using the 41-femur training set. This initial shape model was important for driving automatic image segmentation in the next step. Femur shape was described by a piecewise parametric mesh, which was fitted to manually segmented femur surfaces. For statistical shape modelling, the distribution of fitted mesh nodes should be correspondent across the training set, particularly at well-conserved geometric features. These features were identified on the femur as six well-conserved regions based on Gaussian curvature and cluster analysis. During shape model training, the mesh was fitted to these regions on each training femur, before statistical modelling using PCA. The explicitly defined regions constrained the distribution of nodes in regions of the femur that have been quantified to be correspondent, thereby improving overall correspondence. Average root mean square (RMS) fitting error was quantified to be 0.52 mm, and cross-validation of the shape model in leave-one-out experiments produced an average RMS error of 1.02 mm. In comparison to a non-region based femur shape model, the region-based shape model was more compact and more accurate in approximating seen femurs.

An overview of the shape model design and training process is summarised in figure 6.1. The next section describes how training femurs were partitioned into the six correspondent regions. The method used Gaussian curvature to define regions on individual femurs, and mean-shift clustering to group correspondent regions from across the training set. Section 6.2 presents the femur mesh based on these six correspondent
regions, which was fitted to training set femurs in the shape model training process. Section 6.3 describes this training process, in which region meshes were repeatedly fitted to improve correspondence, before they were assembled for training the full femur shape model. Section 6.4 describes the comparison of the shape model against a non-region-based femur shape model.

### 6.1 Corresponding Regions

The goal of finding correspondent regions across the training set was two-fold: to design a femur mesh composed of these regions, and to fit the mesh to these regions on each training femur.

As explained in chapter 2, a shape model requires a shape description that is correspondent across the training set. Correspondent description can be achieved by partitioning training surfaces into correspondent regions, and optimising correspondence within each region, as demonstrated by (Floater 1997; Zöckler et al. 2000; Lamecker et al. 2002; Horkaew and Yang 2003; Wang et al. 2005). In these works, region partitioning was either done manually, which can be subjective and time consuming, or

---

**Figure 6.1:** Overview of the femur shape model training process.
6.1. CORRESPONDING REGIONS

via intrinsic surface properties, such as curvature, which while reliable, still leaves the problem of having to manually designate correspondent regions between shapes.

In the approach detailed below, the training set of femurs were partitioned automatically using Gaussian curvature. In addition, similar regions from across the training set were clustered and ranked in their correspondence, so that the most correspondent regions could be used in mesh design, and shape model training.

Only surface regions are considered at this stage because its correspondence is easier to achieve through surface landmarks. Volumetric correspondence is complicated by highly variable internal volumetric features. Methods for achieving correspondence of the cancellous bone volume are described in chapter 8.

6.1.1 Methods

Figure 6.2 illustrates the process for identifying correspondent regions on the training set femurs. This process involved the following steps:

1. For each femur triangle surface:
   a) Classify Vertices - label each vertex as having positive or negative local Gaussian curvature;
   b) Partition Regions - group similarly classed vertices through region growing;
2. Cluster Correspondent Regions - group similar regions from all femurs to find those that are the most correspondent.
CHAPTER 6. FEMUR SHAPE MODEL TRAINING

Classifying Vertices

Each vertex was classified by the sign of its local Gaussian curvature, which allowed saddle regions to be discerned from concave or convex regions.

It should be noted that other metrics can also be used to classify vertices, as demonstrated in Kim et al. (2009); Tong and Tang (2005) and Tang and Medioni (2002). In experiments, Gaussian curvature sign gave the most desirable regions, but cortical bone thickness, and a combination of mean and Gaussian curvature were also tried. Their results can be found in appendices 6.C.1 and 6.C.2.

Partition Regions

Figure 6.3: Partitioning regions on the femoral surface. (a) Gaussian curvature estimation at vertices. (b) Vertices classified by Gaussian curvature sign. (c) Preliminary regions produced by region growing similarly classified faces. The $n_r$ largest regions are labelled in non-red colours. (d) Merging of small regions produced the final set of $n_r$ regions (the remaining red region is the smallest of the $n_r$ regions.

Region-growing partitioned femur surfaces into regions with similarly classified vertices, as illustrated in figure 6.3. The method employed was similar to that of Lavoué et al. (2005) and Kim et al. (2009), where regions were grown by the iterative inclusion of faces with similarly-classed vertices (see App. 6.A for details). Region-growing by
faces rather than vertices prevented regions connected by only one edge from being labelled as one region.

Due to noise in Gaussian curvature or triangular mesh imperfections, there were many small spurious regions. These small regions were merged into the $n_R$ largest regions. $n_R$ was set to 10, as an approximate overestimate of the number of desirable regions. This value was validated to be appropriate in a sensitivity analysis (App. 6.B).

Cluster Correspondent Regions

Given the $41 \times n_R$ regions partitioned on the 41 training femurs, similar regions were clustered to find those that were the most correspondent in the training set. For each region, a vector of metrics was calculated, describing the region’s position, shape, and average Gaussian curvature. This allowed similar, and therefore correspondent regions from different femurs to be clustered according to their metric vectors.

The metric vector contained:

- Region surface area, normalised against whole-femur surface area;
- Region centre of mass coordinates, normalised against the bounding box of the femur;
- Region principal moments of inertia;
- Region average Gaussian curvature.

Some of the metrics characterised the position and orientation of the region with respect to the femur. To remove whole-femur translational and rotational variations, all femurs were aligned by their centres of mass and principal axes of inertia before region metric calculation.

Mean-shift clustering ([Cheng, 1995] [Fukunaga and Hostetler, 1975]) was used to analyse the region metric vectors to group correspondent regions. Mean-shift clustering does not require the number of clusters as input, so no assumptions regarding the number of corresponding regions need be made. In addition, it handles clusters of arbitrary shape, and is deterministic. Given a number of data points (metric vectors), mean-shift clustering iteratively updates the position of each point with the mean of all surrounding points weighted by a kernel until the points converge into centres.

A Gaussian kernel was used, in which the kernel width $\beta$ determined the distance
over which data points were clustered, and therefore the number and size of clusters. The value of $\beta$ was chosen to maximise the density and separation of clusters, as measured by the Dunn index \cite{Dunn1974}:

$$D = \min_{1 \leq i \leq n_k} \left\{ \min_{1 \leq j \leq n_k, i \neq j} \frac{d(i, j)}{\max_{1 \leq k \leq n_k} d'(k)} \right\}$$

(6.1)

where $d(i, j)$ is the intercluster distance between clusters $i$ and $j$, measured as the Euclidean distance between their centroids. $d'(k)$ is the intracluster distance measured as the maximum pair-wise distance between points in cluster $k$, and $n_k$ is the total number of clusters.

### 6.1.2 Resulting Regions

Nine major clusters of regions were found, corresponding to major anatomical features (Fig. 6.4). Clusters with fewer than 10 regions were not considered since their regions were found in too few femurs to be reliable. Two measures were used to judge the quality of clusters:

- $D_r(k)$: the average distance of points in cluster $k$ to its centroid.
- $D_{pop}(k)$: the fraction of the training set represented by regions in cluster $k$.

Low $D_r$ suggested a tight cluster, meaning that the regions in the cluster were well conserved, with good correspondence. High $D_{pop}$ suggested that a cluster’s regions were ubiquitous and well represented in the population. Table 6.1 lists these metrics for each of the nine clusters.

The most correspondent regions resembled general anatomic landmarks: the femoral head, greater trochanters and condyles, making them ideal for designing the femur mesh. $D_{pop}$ values in table 6.1 show that these regions corresponded across all femurs. Other important anatomic regions were also present in their own clusters. Figure 6.5 shows these regions on eight randomly selected femurs.

Clusters G to J contained a large number of regions from the femoral shaft. They were highly variable in shape, and did not form tight clusters. This was due to the sensitivity of the Gaussian curvature sign to the shaft shape, which was very close to the transition point between saddle and concave or convex topology. The instability highlighted the effect surface metric had on identifying corresponding regions, specifi-
6.1. CORRESPONDING REGIONS

Figure 6.4: Gaussian curvature sign regions on a typical femur. Labels indicate the cluster each region belonged to. Broad regions corresponding to general anatomical features were consistently identified.

Table 6.1: Region clusters, sorted by increasing cluster mean radius $D_\tau$. Important general anatomic landmarks were identified in clusters in 1 to 4 across all femurs.

<table>
<thead>
<tr>
<th>cluster</th>
<th>$D_\tau$</th>
<th>$D_{pop}$</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.36</td>
<td>1.0</td>
<td>femoral head</td>
</tr>
<tr>
<td>B</td>
<td>0.41</td>
<td>1.0</td>
<td>medial condyle</td>
</tr>
<tr>
<td>C</td>
<td>0.43</td>
<td>1.0</td>
<td>greater trochanter</td>
</tr>
<tr>
<td>D</td>
<td>0.55</td>
<td>1.0</td>
<td>lateral condyle</td>
</tr>
<tr>
<td>E</td>
<td>0.70</td>
<td>0.83</td>
<td>linea aspera</td>
</tr>
<tr>
<td>F</td>
<td>0.72</td>
<td>0.59</td>
<td>lesser trochanter</td>
</tr>
<tr>
<td>G</td>
<td>0.91</td>
<td>0.34</td>
<td>distal shaft</td>
</tr>
<tr>
<td>H</td>
<td>0.98</td>
<td>0.90</td>
<td>proximal shaft</td>
</tr>
<tr>
<td>I</td>
<td>1.20</td>
<td>0.98</td>
<td>shaft fragments</td>
</tr>
</tbody>
</table>
cally, the consequence of a coarse classification metric such as Gaussian curvature sign. In additional experiments, it was found that cortical bone thickness was able to find more consistent boundaries along the femoral shaft (App. 6.C.1). As future work, a combination of surface metrics could be used to identify the best set of regions.

Figure 6.5: Eight randomly chosen femurs labelled with corresponding regions defined by Gaussian curvature sign. The femurs are shown in the anterior view and the posterior view. Regions are labelled and coloured in correspondence to figure 6.4 and region clusters described in table 6.1. Unlabelled regions belonged to clusters that contained less than ten regions. These regions, predominantly along the femoral shaft, were unstable and were replaced by a proximal shaft region and a distal shaft region during femur mesh design and fitting.
6.1.3 Finalising Regions

Based on the clustered regions, each of the 41 training set femurs was partitioned into six regions: the already-partitioned head, greater trochanter, medial condyle, and lateral condyle regions, and the proximal and distal shaft regions (Fig. 6.6).

The proximal and distal shaft regions were bounded by the first four regions, and separated at the mid-shaft. They were introduced because clustered regions beyond the first four were much less stable, or too small in size.

![Image of femur regions](image)

Figure 6.6: The finalised correspondent regions, consisting of the femoral head, greater trochanter, proximal shaft, distal shaft, medial condyle, and lateral condyle.

6.2 Femur Mesh Design

Reference piecewise parametric meshes were designed to efficiently describe the shape of each finalised femur region. The six region meshes were assembled into the ensemble mesh, which described the shape of the whole femur (Fig. 6.7).
The meshes were composed of quadrilateral and triangular elements interpolated by quartic-Lagrange basis, which are described in detail in section 4.2 along with their advantages over other commonly used basis functions. Quartic-Lagrange basis allows elements to have up to three points of inflection in each element-coordinate direction, meaning that a single element can represent a ridge-like, valley-like, convex, or concave feature, all of which are common on the femur. Thus elements were placed over such high-curvature features while element boundaries were placed in regions of low curvature, where possible. Both quadrilateral and triangular elements were used, which allowed greater flexibility in mesh design. For simplicity during assembly, there was a 1-to-1 correspondence between nodes on the boundary of adjacent region meshes (Fig. 6.8). The full ensemble femur mesh was composed of 63 elements and 634 nodes, giving 1902 degrees of freedom, or parameters.
6.3 Shape Model Training

Training the femur shape model using the femur mesh involved two steps, as shown in figure 6.9.
**Region Training:** each region’s reference mesh was fitted to its training set of surfaces (partitioned according to section [6.1.3]), in an iterative fitting-PCA process to improve correspondence;

**Ensemble Training:** the fitted region meshes were then assembled into ensemble meshes, which were fitted a final time, aligned, and used for PCA.

For each region, the fitting-PCA iterations was as follows:

1. For each region surface:
   a) Piecewise Lagrange curves were fitted to the region boundary, or boundaries;
   b) The reference region mesh was fitted to the region, with boundary nodes fixed to the boundary curve, or curves;

2. New reference mesh was calculated as mean fitted mesh;
3. Region shape model was trained using PCA on aligned mesh nodal coordinates;
4. Repeated from step 1, until fitting error converged.

Each region mesh was not necessarily fitted to 41 surfaces. Some clusters of regions did not contain a region partitioned from each of the 41 training femurs, and some regions were excluded due to irregular shape.

The steps above are elaborated upon in the following subsections, before results regarding the fitting accuracy, and the shape model are presented. Fitting methods, objective function, and smoothing functions referenced below are described in detail in section [4.3]

### 6.3.1 Fitting Region Boundary Curves

Quartic-Lagrange curves were fitted to data points (vertices) on the boundaries of training region surfaces to smooth the boundaries, and ensure adjacent region meshes joined when assembled (Fig. [6.10]). The curves contained the same number of elements, and therefore nodes, as the boundaries of their corresponding region meshes.

A curve $B$ was first aligned to boundary vertices $V^b = (v_1^b, \ldots, v_{N_{BD}}^b)$ by finding the optimal translational ($t$), rotational ($r$) and scaling ($s$) transformations

$$ T = (t_x, t_y, t_z, r_x, r_y, r_z, s) $$ (6.2)
6.3. SHAPE MODEL TRAINING

Figure 6.10: Boundary curves. (a) shows a boundary curve (blue, with red nodes), fitted to the boundary of a greater trochanter region, smoothing the jagged boundary. (b) shows a segment of the boundary curve between the greater trochanter and proximal shaft region meshes, with boundary nodes in red. There was a one-to-one correspondence between mesh boundary nodes, and curve nodes, so that the former could be fixed to the latter. This ensured that region boundaries were continuous when assembled.

that minimised the sum of squared distance between the data points their closest points on the curve.

The curve $B$ was transformed according to $T$, and fitted to minimise

$$
\epsilon_{\text{boundary}} = \epsilon_{\text{MD}}(B, V^b) + \epsilon_{\text{DM}}(B, V^b) + \epsilon_s(B) \tag{6.3}
$$

where $\epsilon_{\text{MD}}(B, V^b)$ is the sum of squared distance between points sampled on the curve and their closest data point, while $\epsilon_{\text{DM}}(B, V^b)$ is the inverse: the sum of squared distance between each data point and its closest point on the curve. Details regarding these error terms can be found in section 4.3.1 $\epsilon_s$ is a smoothing term incorpor-
ing Sobolev terms penalising high curvature within elements, and a penalty against mismatched tangent vectors at element boundaries. For more details regarding mesh smoothing, see section 4.3.2. The 2-way distance measure, using both $\epsilon_{MD}$ and $\epsilon_{DM}$, improved the stability of the fit, especially when the boundary was noisy. High smoothing weights were used to encourage smooth curves that capture the general shape of boundaries.

### 6.3.2 Fitting Region Meshes

Each reference region mesh $\Omega^R$ was fitted to the data points (vertices) of training regions $V^R$, with its boundary nodes $x^R_{b_1,\ldots,b_{nB}}$ fixed to the nodes of the fitted boundary curve, or curves $x^B_{1,\ldots,n_B}$.

Fitting was done in four steps:

1. The reference region mesh was aligned to data points using iterative closest point;
2. The aligned mesh was coarsely fitted to the region surface (free-form deformation, or shape model approximation);
3. Mesh boundary nodes were fixed to boundary curve nodes;
4. Mesh was fitted to data points by optimising unconstrained nodes.

Figure 6.11 illustrates these steps on a lateral condyle mesh.

The coarse fit, via free-form deformation (Sec. 4.3.3) in the first iteration, then principal component fitting (Sec. 4.3.3) subsequently, minimised

$$
\epsilon_{\text{coarse}} = \epsilon_{MD}(\Omega^R, V^R) + w_b \sum_{i=1}^{n_B} \|x^R_{b_i} - x^B_i\|^2 \tag{6.4}
$$

where $\epsilon_{MD}(\Omega^R, V^R)$ is the sum of squared distance between points sampled on the mesh and their closest data points (Sec. 4.3.1). The second term is the sum of squared distance between mesh boundary nodes and boundary curve nodes. After this coarse fit, region boundary nodes were fixed to boundary curve nodes by assigning $x^R_{b_i} = x^B_i$ for $i = 1, \ldots, n_B$.

The coarse fit smoothly deformed the mesh to fit large geometric features, giving better convergence for the subsequent nodal mesh fit. Also, it ensured that the boundary nodes were close to their corresponding nodes on the boundary curve, so that there would be minimal change in mesh geometry when boundary nodes were fixed.
6.3. SHAPE MODEL TRAINING

(a) Step 1: training surface with boundary curve
(b) Step 2: aligned reference mesh
(c) Step 3: coarsely fitted mesh, with fixed boundary nodes
(d) Step 4: mesh after nodal fit

Figure 6.11: Region mesh fitting, illustrated using the lateral condyle. The training surface is shown in (a) with its vertices in green. The fitted boundary curve (blue) smoothed the region boundary. The aligned reference mesh is shown in (b), which was coarsely fitted in (c). Mesh nodes are shown in red in (c). Note that those on the boundary are fixed to the curve. Finally, the mesh undergoes a fit of all unfixed nodes (d).

The final nodal mesh fit (Sec. 4.3.3) optimised all nodal coordinates except those at the boundary, minimising

$$
\epsilon_{\text{nodal}} = \epsilon_{\text{MD}}(\Omega^R, V^R) + \epsilon_s(\Omega^R)
$$

(6.5)

where $\epsilon_{\text{MD}}(\Omega^R, V^R)$ is the sum of squared distance between points sampled on the mesh and their closest data points, and $\epsilon_s(\Omega^R)$ is a smoothing terms incorporating Sobolev terms penalising high curvature within elements, and a penalty against mismatched normal vectors at element boundaries.
6.3.3 Iterative Fitting-Training

Each fitted mesh (of a particular region) was aligned, and used to train a region shape model through PCA on aligned nodal coordinates. Nodal coordinates of each mesh were concatenated into a vector $y = (x_1^1, x_1^2, x_1^3, x_2^1, x_2^2, x_2^3, \ldots, x_n^1, x_n^2, x_n^3)$, and assembled into the data matrix $Y = (y_1, \ldots, y_{n_{\text{reg}}})^T$, from which the mean mesh, and principal components of variation were calculated using PCA (Sec. 4.1). Similarly, this was done for the fitted boundary curves.

In the next fitting iteration, the mean mesh and boundary curve were used as the references, and the new shape model was used for coarse fitting: principal component weights, rigid-body translation, and rotation are optimised to generate mesh parameters minimising fitting error. Iterations continued until fitting root mean squared (RMS) errors converged.

6.3.4 Ensemble Mesh Training

After iteratively fitting each femur region, the fitted region meshes for each femur were assembled into ensemble femur meshes. These underwent a final nodal mesh fit to the training set femur surfaces, with all nodal coordinates unconstrained, to minimise errors that may have been caused by fixed boundary nodes. The fitted meshes were finally aligned, and PCA carried out on nodal coordinates to produce the full femur shape model.

6.3.5 Training Results

Results of shape model training on the 41 training set femurs, in terms of iterative fitting-training, fitting accuracy, and resulting principal components of variation, are presented and discussed below.

Effects of Iterative Fitting

Iterative fitting and training improved fitting accuracy, and shape model compactness (Fig. 6.12), which indicated an increase in fitting correspondence. Shape model compactness was measured as the fraction of variance explained by the first five principal
6.3. SHAPE MODEL TRAINING

Figure 6.12: Mesh errors and shape model compactness over fitting-training iterations, for each region mesh. Each line in the left-hand plots shows the shape model approximation error for a particular region. Each line in the centre plot show the nodal fit error for a particular region. The errors were the RMS distance between points sampled on the mesh and their closest data point, averaged across the training set. The errors gradually decreased with the number of iterations. Each line in the right-hand plot shows the fraction of variation explained by the first 5 principal components for a particular region. This value increased rapidly in the first two iterations, and was stable after the third.

components. Approximation errors and compactness stabilised after the third iteration. However, mesh fitting error continued to slowly decrease. Nonetheless, fitted meshes from the third iteration were used to make ensemble meshes. As is shown in the next section, the RMS closest distance between data points and these meshes was about 0.4 mm, which was well below image resolution.

The increase in shape model approximation error over iterations may be due to the influence of an outlier shape on the shape model. PCA is vulnerable to outliers, which can skew principal components away from the true direction of variation. A shape model with a skewed set of components will not be able to fit as well to some, or all shapes, and results in a higher approximation error. Over further iterations, mesh elements may slide (via mesh fitting) into positions better suited to approximating both the outlier and the rest of the shape, resulting in approximation error decreasing
The compactness curves showed a large increase in the first two iterations. The key difference between the first and second iteration was switching from free-form deformation coarse fitting, to shape model approximation. The latter method produced much more correspondent fits, which removed non-shape variations in the fitted meshes, giving more compact shape models. For some regions, compactness increased slightly further in the third iteration, while for others, it decreased slightly. Similarly small fluctuations were seen in further iterations. This could have been due to the shape model approximation propagating residual non-shape variations between components.

Fitting Errors

Across the regions, final fitting error averaged 0.4 mm, and was 0.52 mm for the ensemble (Tab. 6.2). This was less than half of the voxel resolution of the VIFM images, so the fitted meshes were considered faithful representations of manually segmented femur geometry.

Table 6.2: Average fitted RMS errors for each of each of the region meshes, and the ensemble femur mesh.

<table>
<thead>
<tr>
<th>Region</th>
<th>Average RMSE (1 S.D.) (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>0.30 (0.069)</td>
</tr>
<tr>
<td>Greater Trochanter</td>
<td>0.42 (0.079)</td>
</tr>
<tr>
<td>Proximal Shaft</td>
<td>0.42 (0.062)</td>
</tr>
<tr>
<td>Distal Shaft</td>
<td>0.40 (0.045)</td>
</tr>
<tr>
<td>Lateral Condyle</td>
<td>0.43 (0.087)</td>
</tr>
<tr>
<td>Medial Condyle</td>
<td>0.44 (0.066)</td>
</tr>
<tr>
<td>Ensemble</td>
<td>0.52 (0.091)</td>
</tr>
</tbody>
</table>

Components of Variation

The first four principal components of the femur shape model accounted for over 95% of total variation, and reflected obvious changes in shape and size (Fig. 6.13).

The first component was dominated by size variation, plus a small increase in the femoral neck angle with size. The second component accounted for variation in anteverision angle, that is, twisting of the femur along the shaft axis. The third component
accounted for femoral neck angle, inversely related to widths throughout the femur, while the fourth component reflected an increase in the neck angle, correlated with an increase in the sizes of the proximal and distal femur.

**Figure 6.13:** Principal components of the ensemble femur shape model. (a) to (d) show shape variation along the first four principal components, at -2 and +2 standard deviations.

### 6.4 Comparison to Non-regional Shape Model

The hypothesis of the region-based shape model was that a higher quality shape model resulted from enforcing correspondence in individual regions known to correspond. This hypothesis was tested by comparing this *regional* shape model, to a *non-regional*
shape model. Training of the non-regional shape model is described next, followed by comparison metrics, and finally the results and discussion.

6.4.1 Non-Regional Shape Model

The non-regional shape model was trained on the same 41-femur training set. The average ensemble mesh was picked as the non-regional reference mesh. This reference mesh was fitted to the femur surfaces in three steps:

1. ICP alignment,
2. Free-form deformation coarse fit,
3. Nodal fit.

All steps were performed on the whole femur, with no partitioning of regions. Fitted meshes were aligned and used to train a shape model through PCA.

6.4.2 Comparisons

Regional and non-regional shape models were compared in terms of

- **Fitting Accuracy**: RMS error of mesh fitting during the training phase.
- **Compactness**: number of principal components needed to describe a fixed percentage of variation.
- **Generality**: RMS error of shape model approximation in leave-one-out cross-validation.

The leave-one-out cross-validation for shape model generality was performed as follows. For each of the fitted meshes, a shape model was trained using all other fitted meshes, and the shape model was used to approximate the left out mesh. Approximation error was measured in two ways:

- **Closest Point Error**: RMS distance of points on the left-out mesh to their closest points in the approximated mesh. (9000 evenly distributed points). This measured the ability of the shape model to accurately approximate unseen shapes, in terms of geometric accuracy.
- **Corresponding Point Error** RMS distance between corresponding points on the left-out mesh, and the approximated mesh. (9000 evenly distributed points).
6.4. COMPARISON TO NON-REGIONAL SHAPE MODEL

This measured the ability of the shape model to describe an unseen shape in a correspondent manner to its training set. It also gave a measure of the level correspondence in the training set.

Leave-one-out cross validation was performed using 5, 10, and 20 principal components for approximation.

6.4.3 Results and Discussion

Fitting Accuracy

Across the training set, fitting accuracy for the non-regional shape model was 0.61 mm, with 0.21 mm standard deviation. This was significantly greater than that of the regional shape model (0.52 mm ± 0.09 mm). This shows that fitting region meshes, then assembling, gave a more accurate description of femur geometry. The reduction in error likely resulted from the region-specific shape model approximations used during region fitting iterations. The region-specific approximations were more accurate in each region than the whole-femur free-form deformation of the non-regional model, which lacked the degrees of freedom to achieve a closer fit. The closer coarse fits provided by the shape models led to more accurate nodal fits. Even if the degrees of freedom was increased for the free-form deformation, or free-form deformation was performed iteratively for each region, it would still lack the real shape data-defined constraints of the shape models, and be much more computationally expensive.

Shape Model Compactness

The regional shape model was more compact in the absence of size variations (Fig. 6.14). With size variation included, it dominated the first component, and there was little difference between the two methods. However, when size variation was removed, a difference was obvious. The regional shape model described a greater proportion of variations in the first three components. This meant that the regional shape description and training process was capable of producing more compact shape models.
Shape Model Generality

The regional shape model produced lower leave-one-out approximation errors when a small number of components was used (Tab. 6.3). There was no significant difference when more components were used, but error did decrease for both shape models, from about 1.5 mm, to 1.0 mm. Note that these errors were calculated against left out meshes, not the manually segmented surfaces to which the meshes were fitted. In the latter case, the non-regional shape model’s error would be higher, owing to its poorer fitting accuracy reported above (Sec. 6.4.3).

The regional shape model also produce lower corresponding point errors, with fewer principal components. In addition, error standard deviation was about half of that of the non-regional shape model. This suggested that the regional shape model was much more correspondent in parameterising the femur surface. Lower error with fewer components again suggested that the regional shape model was more compact, with
Table 6.3: Leave-one-out approximation errors of the regional and non-regional shape models.

<table>
<thead>
<tr>
<th>components</th>
<th>non-regional</th>
<th>regional</th>
<th>non-regional</th>
<th>regional</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.57 (0.49)</td>
<td>1.42 (0.45)</td>
<td>3.37 (1.59)</td>
<td>3.16 (0.79)</td>
</tr>
<tr>
<td>10</td>
<td>1.19 (0.33)</td>
<td>1.18 (0.34)</td>
<td>3.03 (1.82)</td>
<td>2.93 (0.70)</td>
</tr>
<tr>
<td>20</td>
<td>1.03 (0.28)</td>
<td>1.02 (0.25)</td>
<td>2.83 (1.70)</td>
<td>2.90 (0.72)</td>
</tr>
</tbody>
</table>

more meaningful variations captured in the first few components.

The difference between the regional and non-regional errors was smaller than expected. This could be due to the non-regional reference mesh. The reference mesh was the average mesh produced from the regional shape model, which was an optimised average of the all training set femurs. It would have fitted more consistently to each surface, and gave a better shape model, than if a particular training surface was used as the reference. Thus the difference in error from the presented experiment was likely to be conservative. If an optimised average mesh was not available, the global-fit shape model would produce worse results in the leave-one-out cross-validation.

6.5 Conclusions and Future Work

This chapter has covered the training of the initial femur statistical shape model, on a training set of 41 femurs. Training femurs were automatically partitioned into regions using Gaussian curvature, and mean-shift clustering was used to identify six major regions, into which the training set femurs were divided. A reference piecewise Lagrange parametric mesh for each region was created, and fitted to its corresponding region on training femurs. Fitting was iterated using region shape model approximations to improve regional correspondence. Fitted region meshes were then assembled into full femur meshes to train the full femur shape model using PCA. Compared to a non-regional femur shape model, the regional shape model was more compact, and more correspondent, producing lower errors in leave-one-out cross-validation.

The correspondent region partitioning method produced regions given a population
of femur. However, at this stage, the method does not find the corresponding regions on a new femur. This may be useful when the training set expands, and new femurs need to be added to the shape model. This function can be added by training a statistical classifier to assign the regions on a new surface to existing clusters of regions.

A statistical classifier could also improve the current region partitioning process. The classifier can be trained using ideal regions (e.g., manually defined), and used to guide the partitioning process. This could improve robustness by avoiding the problem of a region being split into multiple fragments (e.g., in the femoral shaft).

Improving the partitioning process should produce more consistent clusters of corresponding regions, which should reduce the clustering method’s sensitivity to its kernel width parameter ($\beta$). This would make it more feasible to automatically select the optimal partitioning and clustering parameters.

The region meshes corresponded to anatomically relevant regions, and their individual shape models produced components of variation that may be interesting from a functional and clinical point of view. Thus regional shape models warrant further investigation once the full VIFM dataset has been segmented and meshed.

Another future goal is to relate the shape models of different regions, using methods such as canonical correlation analysis ([Hardoon et al., 2004]), or partial least-squares regression ([Abdi, 2003]). The field of view of femur images can be restricted, for example to the knee, or hip joint. With related regional shape models, the geometry of other femur regions can be predicted from the limited geometry in these images.
6.A Region Growing and Merging Algorithm

Region Growing

The region-growing procedure aims to collect connected vertices of the same classification. The method we employ is similar to that of Lavoué et al. (2005) and Kim et al. (2009), where regions are grown by the inclusion of faces. Region-growing by faces rather than vertices prevents regions connected by only one edge from being labelled as one region.

First, all triangular faces with all three vertices of the same classification are collected in the set $t_{\text{seeds}}$. These are candidates for seeding regions. A seed face $t_{\text{seed}}$ in $t_{\text{seeds}}$ is picked at random to start growing the first region $r_1$. The region’s classification $g_{r_1}$ is that of its vertices. At this stage, $r_1$ consists of the face $t_{\text{seed}}$ and its vertices.

The region is grown iteratively. As shown in figure 6.15 in each iteration, for each face $t_{i}^{\text{ext}}$ immediately exterior to $r_1$ (initially, the three faces neighbouring $t_{\text{seed}}$), its vertex $v_{i}^{\text{ext}}$ exterior to $r_1$ is tested for its classification $g_{v_{i}^{\text{ext}}}$. If $g_{v_{i}^{\text{ext}}}$ matches $g_{r_1}$, then $v_{i}^{\text{ext}}$ and $t_{i}^{\text{ext}}$ are added to $r_1$. Once all $t_{i}^{\text{ext}}$ have been examined and all eligible faces and vertices added to $r_1$, the next iteration begins with the expanded $r_1$.

When no new faces can be added to a region, a new unlabelled $t_{\text{seed}}$ is randomly selected from $t_{\text{seeds}}$ to grow a new region ($r_2, r_3, ...$) following the procedure above until all faces in $t_{\text{seeds}}$ have been assigned to a region. All remaining free faces and associated vertices are assigned to their neighbouring region sharing the most number of edges.

Region Merging

Due to noise in the vertex metric or triangular mesh imperfections, there may be many small spurious regions. Let the $n_R$ largest regions belong to the set $R$, and all smaller regions belong to the set $r$. In the region refinement step, all regions in $r$ are merged into the regions in $R$, to give $n_R$ regions. $n_R$ is one of the user-defined parameters of our method.

The first step of region refinement merges neighbouring regions in $r$. Regions in $r$ are ranked by size in terms of number of vertices. Starting with the smallest, each region in $r$ with neighbouring regions $r^*$ (a subset of $r$) is merged with the region in $r^*$
Figure 6.15: The region growing process. Vertex and triangle colours indicate their classification, pink triangles are those immediately external to a region. In figure (a) a new region is started at the face \( p_{\text{seed}} \), which has 3 vertices of the same class. In the first iteration, its three neighbours coloured in pink are checked for the classification of their external vertices. \( t_{1}^{\text{ext}} \) and \( t_{3}^{\text{ext}} \) are added to the region, \( t_{1}^{\text{ext}} \) is not, and is ignored. This results in the region shown in figure (b) at the start of the second iteration. Here, since \( t_{2}^{\text{ext}} \) and \( t_{3}^{\text{ext}} \) will be added to the region, so will the triangle between them, giving the starting region for iteration 3, shown in figure (c).
with which it shares the largest number of vertices, until no more merges are possible. After this, all regions in \( r \) will either be entirely surrounded by a single large \( R \) region, or be situated between two or more large \( R \) regions. Those in the former case are merged into the large region they are surround by. Those in the latter case are merged with the neighbouring large region with which they shared the most vertices. Now all vertices and faces are assigned to one of the \( n_R \) regions in \( R \).

### 6.B Correspondent Region Sensitivity Analysis

![Dunn's Index](Dunn's Index.png)

**Figure 6.16:** Dunn’s Index for regions cluster at a range of \( \beta \) and \( n_R \) values, for regions defined by vertex Gaussian curvature sign. Correspondent regions clustered using parameters at peaks are inspected to determine the optimal \( n_R \) and \( \beta \).

Regions produced by Gaussian curvature sign seemed ideal for designing the femur mesh, and training the shape model. However, there was the question of the stability of these regions. Besides the choice of vertex metric, the other free parameters in the method were the number of regions to segment per surface \( n_R \), and the clustering kernel width \( \beta \).
Region partitioning and clustering was run at a range of $n_R$ and $\beta$ values for Gaussian curvature sign classification. The Dunn’s Indices of all combinations are plotted in figure 6.16, where warmer colours represent tighter and better separated clusters, i.e. well-conserved regions that were distinct from regions in other clusters. Regions clustered at significant peaks in this parameter space were visually inspected.

The parameters responsible for the regions presented in section 6.1.2 were $n_R = 10$, and $\beta = 0.85$, which corresponded to the highest peak upper-centre of figure 6.16. Clusters at higher $\beta$ values became too large, and contain dissimilar regions. At low $\beta$ values, many small clusters were produced that contain similar regions between them. At low $n_R$, clusters with high Dunn’s indices were produced. However, the regions had a tendency to be too broad. For example, the greater trochanter and the femoral shaft would be a single region. At high $n_R$ values, regions partitioned on surfaces became smaller, and more unstable in their shape. This made it harder to find regions across surfaces that are similar.

In the area around the $n_R = 10$, $\beta = 0.85$ peak, there is a quick drop-off in the Dunn’s Index, and cluster quality. So overall, it appears that while the regions produced are sensitive to $n_R$ and $\beta$, their optimum values are easy to identify through the process used in this sensitivity analysis.

6.C Addition Correspondent Region Experiments

6.C.1 Cortical Thickness

Table 6.4: Cortical thickness region clusters, sorted by increasing cluster mean radius $D_r$. All three clusters can be considered to hold corresponding regions due to the consistency and ubiquity of the regions across the population.

<table>
<thead>
<tr>
<th>cluster</th>
<th>$D_r$</th>
<th>$D_{pop}$</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.51</td>
<td>0.98</td>
<td>proximal femur</td>
</tr>
<tr>
<td>B</td>
<td>0.80</td>
<td>0.98</td>
<td>distal shaft</td>
</tr>
<tr>
<td>C</td>
<td>1.05</td>
<td>0.98</td>
<td>femoral shaft</td>
</tr>
</tbody>
</table>

Table 6.4 lists the clusters of regions identified using cortical thickness. The three clusters contained the proximal femur, the distal femur, and the femoral shaft regions.
Figure 6.17: Cortical thickness regions on a typical femur. Labels indicate the cluster each region belongs to. The proximal femur, shaft, and distal femur were consistently identified.

(Fig. 6.17) from all femurs except one. In the one exceptional case, the femur had abnormally thin cortical bone on the lateral and medial sides of the femoral shaft, which were identified as separate regions. Clustering correctly excluded these abnormal regions from the three clusters.

According to Ruedi and Buckley (2007), and discussions with an orthopaedic surgeon, the proximal femur-shaft boundary should be just distal to the lesser trochanter. For the distal femur, the border should be just proximal of the adductor tubercle (Fig. 6.18). However, there are no precise definitions for these boundaries.

The boundary given by our method closely matched the above definition. Cortical thickness varied circumferentially around the femoral shaft. Therefore, so did the position of the region boundary, since regions were partitioned according to cortical thickness. As expected, the boundary was less stable in the inferior femoral neck, where
the cortex is thicker than in the rest of the proximal femur.

The distal femur-shaft boundary given by our method was more proximal along the femoral shaft than the clinical definition. This was due to K-Means clustering deciding the critical thickness value that separated the regions. However, the position of the boundary was consistent across the population.

6.C.2 Mean and Gaussian Curvature

Table 6.5: Mean and Gaussian curvature region clusters, sorted by increasing cluster mean radius $D_r$.

<table>
<thead>
<tr>
<th>cluster</th>
<th>$D_r$</th>
<th>$D_{pop}$</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.31</td>
<td>0.44</td>
<td>lateral condyle ridge</td>
</tr>
<tr>
<td>B</td>
<td>0.36</td>
<td>0.59</td>
<td>lateral greater trochanter</td>
</tr>
<tr>
<td>C</td>
<td>0.54</td>
<td>0.95</td>
<td>lesser trochanter</td>
</tr>
<tr>
<td>D</td>
<td>0.54</td>
<td>0.95</td>
<td>proximal greater trochanter</td>
</tr>
<tr>
<td>E</td>
<td>0.62</td>
<td>0.95</td>
<td>medial condyle ridge</td>
</tr>
<tr>
<td>F</td>
<td>0.67</td>
<td>0.90</td>
<td>linea aspera</td>
</tr>
<tr>
<td>G</td>
<td>0.73</td>
<td>0.95</td>
<td>femoral head fragments</td>
</tr>
<tr>
<td>H</td>
<td>0.83</td>
<td>0.95</td>
<td>lateral condyle ridge</td>
</tr>
<tr>
<td>I</td>
<td>0.94</td>
<td>0.90</td>
<td>shaft ridges</td>
</tr>
<tr>
<td>J</td>
<td>1.07</td>
<td>0.76</td>
<td>whole femur</td>
</tr>
</tbody>
</table>
Figure 6.19: Mean and Gaussian curvature regions on a typical femur. Labels indicate the cluster each region belongs to. Higher curvature ridges were consistently identified.

Regions identified using Gaussian and mean curvature tended to be sharp ridges and protrusions (Fig. 6.19). All main regions of this nature were represented in 10 clusters listed in table 6.5. The proximal ridge of the greater trochanter, the lesser trochanter, the ridges on the medial and lateral condyles were identified on 95% of the tested population.

Some clusters did not contain a region from every femoral surface. In some cases, due to variability between femurs, a region was not prominent enough to be distinguished from its surroundings in the vertex classification step, and therefore was not identified. For example, clusters B captured the protrusion on the lateral side of the greater trochanter from only about 60% of the 41 femurs. However, this was expected as shape of the greater trochanter varies greatly between femurs (Hislop-Jambrich 2010). This protrusion was not partitioned on femurs with relatively smooth greater trochanters.
While this may suggest poor correspondence across the population, the cluster’s low $D_r$ scores (cluster width) suggests that its regions were very similar. Therefore, this regions should still be considered as good correspondent region, and parameters for vertex classification can be adjusted to improve the sensitivity.

In other cases, a single region was sometimes identified as two. For example, the lateral condyle ridge was found in two clusters. Cluster A contained the anterior portion of the ridge only, while cluster H contained the ridge in its entirety, and its posterior portion. Adjustment of vertex classification parameters may improve region partitioning robustness, leading to better quality clusters in terms of $D_r$ and $D_{pop}$. Alternatively, the use of other clustering methods, such as evolving mean-shift clustering (Zhao et al., 2010), could be investigated.

The ridge-like regions serve well for enforcing model correspondence when training statistical shape models. For example, in the classical approach where a template model is fitted to training shapes, higher fitting weights can be assigned to these regions. This improves upon methods where landmarks are chosen manually, are restricted to points, and affected by the subjectivity of the operator. Instead of enforcing the correspondence at sparse points, correspondence can be directed, in the case of this example, using the shape of natural geometric boundaries.
Chapter 7

Automatic Cortex Segmentation and Meshing

The femoral cortex is a crucial part of femur morphology. It dictates femoral geometry, function, and strength. Kinematic, finite-element, and morphometric analysis of the femur all depend on accurate cortex geometry. This chapter presents a fully automated method for segmenting and meshing cortical surfaces from clinical CT images. Active shape modelling, and cortical thickness mapping were combined, and adapted for the piece-wise parametric femur mesh. Mesh error was validated to be under 0.93 mm RMS. Testing on the full VIFM dataset produced good quality meshes from 83% of scans.

The work of this chapter was published in the proceedings of Mesh Processing in Medical Image Analysis 2012 (Zhang et al., 2012).

7.1 Introduction

The femoral cortex (Fig. 7.1) is a vital consideration in nearly all femur-related analyses. From an anthropological and forensic perspective, cortical morphology varies with factors such as age (Bertelsen et al., 1995), height (Looker et al., 2001), and ethnicity (Peacock et al., 2009). In terms of predicting hip fracture risk, cortical geometry and thickness distribution have been shown to be important factors (Holzer et al., 2009; Pulkkinen et al., 2004; Mayhew et al., 2005). Cortical geometry and thickness are also
Figure 7.1: Coronal slice through the proximal femur, with the cortex outlined in red. Loads on the femurs are mainly transmitted through the cortex, which is the main determinant of femur strength. Cortical thickness varies widely over the femur, and can be smaller than image resolution in clinical CT scans.

important for the accuracy of finite-element models of femur mechanics [Zdero et al., 2010]. These are all relevant applications for a statistical model of femur morphology, so automatically segmenting and modelling the femoral cortex is an integral part of this project.

The prevalence of X-ray computed tomography (CT) means that a wealth of 3-D femoral cortex data exists in clinical CT images. However, subjective manual segmentation of the images is prone to error, and, when large datasets are required, the task becomes prohibitively laborious. It is thus beneficial to automate the accurate segmentation and modelling of femoral cortex from clinical CT images.

Active Shape Modelling (ASM) [Cootes et al., 1995], and similar methods, have been used for the automatic segmentation of a wide range of anatomical structures [Heimann and Meinzer, 2009]. ASM uses a statistical model of object shape, and object appearance normal to its surface, to robustly segment the object’s surface from an image. At the same time, segmented points on the surface are correspondent across
7.2. SEGMENTATION AND MESHING PROCESS

subjects, which enable statistical analysis on a population of segmentations. ASM provides a way to locate the femoral surface before more accurate cortex segmentation.

In clinical CT images, the accuracy of cortical surface estimation is limited by the image resolution. Simple methods such as thresholding become inaccurate, or fail altogether when thin cortex is blurred by partial volume effects (Dougherty and Newman [1999]). Recently, Treece et al. (2010) proposes modelling the appearance of the cortex using a function that accounts for imaging characteristics. The model is able to predict the position of the inner and outer cortical surface to sub-pixel accuracy. Their method, Cortical Thickness Mapping (CTM), gives significantly more accurate thickness estimates than previous methods.

Both ASM and CTM require an underlying surface description. Typically, triangular surfaces are used. This chapter will describe how these methods have been adapted for the piecewise parametric mesh of the femur. The cortex segmentation and meshing process is presented in the next section. Validation experiments and results are presented in section 7.3. Following that, results from segmenting and meshing the full VIFM dataset is presented in section 7.4. Finally, limitations and scope for future work are discussed in section 7.5.

7.2 Segmentation and Meshing Process

The segmentation and meshing process begins with the mean femur mesh (from the femur shape model, chapter 6) which is customized to represent the inner and outer cortical surfaces in a CT image in three main steps (Fig. 7.2):

1. An active shape model deforms the generic mesh to approximate the in-image femoral surface;
2. CTM is carried out on the ASM mesh, to estimate cortical thickness and segment the inner and outer cortical surfaces;
3. Data from CTM is used to further deform the ASM mesh to represent the inner and outer cortical surfaces.

The following section describes the femur active shape model, and its use in the segmentation process. Section 7.2.2 describes how CTM is implemented on the femur
162

CHAPTER 7. CORTEX MESHING

Figure 7.2: The femur mesh at key steps during cortical surface segmentation. The mean mesh is first deformed by ASM segmentation, then used for CTM, and finally deformed further into the inner and outer cortical surface meshes.

The focus will be on the application of ASM and CTM to femoral cortex segmentation. Details regarding the theory of these methods can be found in sections 4.4 and 4.5.

7.2.1 ASM Segmentation

The ASM segmentation procedure is composed of an initialisation step, then surface prediction and mesh fitting iterations at two resolutions, as illustrated in figure 7.3. The femoral shaft is located, and used to initialise femur mesh position and orientation. Iterative ASM segmentation follows, where surface prediction, error correction, and mesh fitting are repeated until convergence. Iterative segmentation is first performed at a coarse resolution, then at a finer resolution. Finally, the ASM-predicted mesh is fitted to the segmented data in a conventional nodal mesh fit (Sec. 4.3.3).

The ASM algorithm used is based on Cootes et al. (2001), adapted for piecewise-parametric meshes. The key difference is that models of image appearance are based at fixed element coordinates (material points) within each mesh element, rather than
the vertices of a triangulated mesh.

**Femur ASM Training**

The femur ASM was trained using the meshes and CT images of 41 femurs (same as that used to train the initial shape model). Statistical models of image texture normal to the femoral surface (local appearance models) were trained at $Q_{ASM}^{\text{ASM}} = 4230$ points, distributed in ten by ten grids within each mesh element. See section 4.4.1 for details regarding the training process.

The models were trained at two different resolutions to allow hierarchical ASM segmentation. The low resolution model was trained with profiles 60 samples long, at a resolution of 2 mm. The high resolution model was trained with profiles 40 samples
long, at a resolution of 1 mm.
Initialisation and Segmentation

Segmentation involves producing a femur mesh \( \Omega \) that approximates the in-image femur surface. \( \Omega \) is generated from the femur shape model with principal component weights \( \mathbf{a} \), then transformed by a rigid-body translation \( \mathbf{t} \) and rotation \( \mathbf{r} \). \( \mathbf{a} \) contains the weights of the first five principal components, which covered over 95\% of training set variation.

Reliable ASM segmentation requires close initial placement of the shape model mesh. To this end, the initial femur mesh is aligned to the femoral shaft. The shaft is first located by searching for large, isolated cylindrical structures in the CT scan. The shaft surface is then segmented by thresholding and simple edge-detection (this is possible due to the lack of nearby bony structures), to give a data-cloud \( \mathbf{D}^{\text{shaft}} \). Details of shaft segmentation is given in appendix 7.A.

ASM segmentation is initialised by aligning the mean femur mesh to \( \mathbf{D}^{\text{shaft}} \). First, \( \mathbf{t} \) is set to the vector from the centre of mass of the mean mesh, to that of \( \mathbf{D}^{\text{shaft}} \), while \( \mathbf{r} \) is a pre-set rotation (Fig. 7.4b). Second, \( \Omega \) is fitted to \( \mathbf{D}^{\text{shaft}} \) by optimising \( \mathbf{a} \), \( \mathbf{t} \), and \( \mathbf{r} \) (Fig. 7.4c).

After initialisation, iterative ASM segmentation proceeds as detailed in section 4.4.2:

1. Search for the femur surface at \( Q^{\text{ASM}} \) material points, normal to the mesh;
2. Identify and correct outliers in the predicted surface points;
3. Fit the mesh to the surface points by optimising \( \mathbf{a} \), \( \mathbf{t} \), and \( \mathbf{r} \);
4. Go to step 1 until mesh is stable between iterations, or maximum number of iterations is reached.

The error correction step significantly improves surface prediction accuracy, and takes advantage of the piecewise construction of the femur mesh (Fig. 7.5). After step 1 above, for each mesh element, the median and standard deviation of the predicted surface points are calculated. Points positioned greater than one standard deviation away from the median are flagged as outliers. These points are then repositioned at a possible surface closest to the median position.

Segmentation is first run using the low-resolution appearance model with long profiles. This allows far-away surface matches to be found while avoiding false small scale features, e.g. adjacent bones at joints, or other parts of the femur itself. After the
low resolution segmentation terminates, segmentation is run with the high resolution local appearance model with shorter profiles, to refine the segmentation. Once this terminates, the outputs are a point cloud $D^{ASM}$ of the final ASM-predicted surface, and the femur shape model’s best fit (Fig. 7.4d).

**Final Mesh Fit**

The shape model cannot accommodate variations not represented by its training set. A final conventional mesh fit to $D^{ASM}$ (Fig. 7.4e) is thus performed to capture the finer details of the segmented surface. The fit optimises all mesh node coordinates, subject to Sobolev smoothing. The fitted mesh $Ω^{ASM}$ is now close enough to the cortex for CTM to find the cortical surfaces.

### 7.2.2 Cortical Thickness Mapping on Mesh

Following ASM segmentation, cortical thickness mapping (CTM) is used to obtain a more accurate segmentation of the inner and outer cortical surfaces, and measure cortical thickness. CTM is performed on a 15 by 15 grid within each element of $Ω^{ASM}$, giving a set of $Q^{CTM} = 9420$ material points. At each point, the cortex profile model $y_{blur}$ is fitted to an image sample $\hat{y}$ normal to the mesh, as illustrated in figure 7.6. For details on the profile model and aspects about fitting, see section 4.5.
7.2. SEGMENTATION AND MESHING PROCESS

Figure 7.6: Cortical thickness mapping. An image sample across the cortex (sample) is approximated by $y_{\text{blur}}$, which is produced from $y$ with cortical surface positions $c_0$ and $c_1$.

Fitted parameters are:

- $c_0$ - inner cortical surface position,
- $c_1$ - outer cortical surface position,
- $y_0$ - internal image intensity,
- $y_2$ - external image intensity,
- $\sigma$ - image point spread function width.

Cortical image intensity $y_1$ is fixed to the highest intensity in the femoral shaft, where the thick cortex is unattenuated by partial-volume effects. Initial positions of the inner and outer cortical surface ($c_0$, $c_1$) along each CTM profile are automatically placed at the positions of maximum and minimum gradient, respectively. Due to the proximity of the pelvis, this is not done for profiles sampled from the femoral head, where initial $c_0 = -4.0$ mm and $c_1 = 2.0$ mm (inside, and outside the mesh, respectively). Extra-cortical tissue intensities $y_0$ and $y_2$ are both set to 0 HU, and $\sigma$ to 1.6. The length and resolution of samples are set following the advice of Treece et al. (2010) and our own experimentation. In the proximal and distal femur regions, sample length is 18 mm, while in the femoral shaft, sample length is 24 mm. Profile resolution is 0.1 mm.
Performing CTM on all $Q^{\text{CTM}}$ material points produces

- $D^\text{in} = \{d^\text{in}_{1, \ldots, Q^{\text{CTM}}} \}$ - point cloud of the inner cortical surface,
- $D^\text{out} = \{d^\text{out}_{1, \ldots, Q^{\text{CTM}}} \}$ - point cloud of the outer cortical surface,
- $t = \{t_{1, \ldots, Q^{\text{CTM}}} \}$ - cortical thickness estimates.

These are then used to customize $\Omega^{\text{ASM}}$ to produce inner and outer cortical surface meshes. When CTM fails at a material point (determined by thickness threshold and fitting error), its thickness value and cortical surface estimates are masked from subsequent processing. In the femoral head, the very thin cortex causes a high incidence of failure cases. To avoid not having enough data points for well-conditioned fitting (see Sec. 7.2.3), failed thickness estimates are set to 0.1 mm, which is an estimate of minimal cortical thickness in the femoral head.

### 7.2.3 Inner and Outer Cortical Surfaces

![Figure 7.7: Mesh customisation using CTM data. (a) and (b) show, respectively, point clouds of the outer ($D^{\text{out}}$) and inner ($D^{\text{in}}$) cortical surface produced by CTM. (c) shows the cortical thickness field $\Psi$ fitted to thickness estimates from CTM. These three dataset are used to customise the ASM-segmentation mesh to the outer and inner surfaces (d).](image)

The inner and outer cortical surface meshes, $\Omega^{\text{in}}$ and $\Omega^{\text{out}}$ respectively, are both produced by customizing $\Omega^{\text{ASM}}$ according to $D^{\text{in}}$, $D^{\text{out}}$ and $t$ (Fig. 7.7). $\Omega^{\text{out}}$ is created...
by fitting $\Omega^{ASM}$ to $D^{out}$. However, due to noise and fitting errors, simply fitting $\Omega^{ASM}$ to $D^{in}$ for the inner cortical mesh may result in the thickness between $\Omega^{in}$ and $\Omega^{out}$ not accurately reflecting CTM estimates.

Over the proximal and distal regions of the femur, the cortex is thin enough to assume that the inner and outer cortical surfaces have coincident surface normals. Because of this, and the fact that $\Omega^{out}$ and $\Omega^{in}$ share the same mesh topology, the proximal and distal regions of $\Omega^{in}$ are created by offsetting its nodes inwards from their positions in $\Omega^{out}$ (Fig. 7.8). The direction of the offset is along the surface normal evaluated at the nodes.

The offset for each node is determined by the thickness field $\Psi$, a scalar field interpolated over the femur mesh (Fig. 7.7c). $\Psi$ is fitted to thickness measurements by minimising the sum of squared differences between the values of $\Psi$ and $t$ at each material point CTM was performed. Because $\Psi$ shares the same mesh topology with $\Omega^{in}$ and $\Omega^{out}$, the fitted thickness value at a node is equivalent to the inward offset of that node in $\Omega^{in}$ (in the proximal and distal regions).

In the femoral shaft, the thick cortex and prominent ridges mean that $\Omega^{in}$ and $\Omega^{out}$ normals do not coincide at many places. Displacing nodes in the same manner as for the proximal and distal regions results in self-intersecting meshes. Instead, the shaft

\textbf{Figure 7.8:} Cortical mesh construction by inward projection. The proximal and distal femur regions of the inner cortical mesh is created by projecting the outer mesh inwards according to estimated cortical thickness.
region of $\Omega^\text{in}$ is set to be cylindrical in shape. Then, shaft elements are fitted to data points in the shaft region of $D^\text{in}$.

### 7.3 Validation Experiment

The accuracy of femoral cortex segmentation and meshing was validated in a leave-one-out experiment, and in comparison to an established semi-automated method. First, the leave-one-out experiment tested the ability of the ASM to locate the femur, and provide a close enough estimate of the femoral surface for CTM. Second, the accuracy of cortical surface meshes was validated against the Stradwin software (Fig. 7.9). Stradwin is developed by the authors of CTM (Treece et al., 2010), and performs CTM on manually segmented surfaces. Stradwin has been validated on high-resolution peripheral CT images of the proximal femur, giving a mean thickness error of 0.01 mm ± 0.58 mm.

![Figure 7.9: User interface of Stradwin. In Stradwin, CTM is performed on manually segmented surfaces, shown in the two upper panes. The lower panel shows the CTM profile (red) fitted to the sampled profile (blue).](image)

Errors calculated were the root mean square (RMS), and standard deviation, of the
closest distance between ground-truth data points and the mesh surface.

7.3.1 Leave-One-Out

The leave-one-out ASM segmentation validation was carried out using the 41 training set femurs. For each image, an ASM was trained using all other images. This ASM was used to segment the left femur in the image, and the segmented mesh was compared against the manual segmentation, which was the ground-truth.

Of the 41 femurs left out and segmented, ASM-segmentation for one did not converge, and this femur was omitted from analysis. The proximal femur in this image was poorly predicted by the ASM. Unusual femoral shape was likely the cause.

Disregarding the single failed segmentation described above, the mean RMS error for the segmentations was 1.02 mm, with a standard deviation of 0.74 mm. This result shows that, except in extreme cases, ASM segmentation found the femoral surfaces to voxel-resolution accuracy. Given that the length of CTM profiles are 20 mm in length, the accuracy of ASM is sufficient in providing an initial surface estimate for the subsequent CTM.

7.3.2 Cortical Surface Meshes

Seventeen CT scans not in the ASM training set were used to validate the accuracy of the cortical surface meshes produced by the full CT image-to-mesh process. The resulting inner and outer cortical meshes ($\Omega^\text{in}$ and $\Omega^\text{out}$, respectively) were compared against inner and outer polygon meshes produced by Stradwin ($\Omega^{\text{in}*}$ and $\Omega^{\text{out}*}$, respectively), which were considered as ground-truths.

In addition, mesh thickness was compared against Stradwin thickness estimates associated with each vertex on $\Omega^{\text{out}*}$. Mesh thickness was calculated as the Euclidean distance between $\Omega^\text{in}$ and $\Omega^\text{out}$ at a common material point. For each vertex in $\Omega^{\text{out}*}$, the difference between its thickness value and the mesh thickness of its closest point on $\Omega^\text{out}$ was calculated. Note the closest point on $\Omega^\text{out}$ was not restricted to the CTM material points. The RMS and standard deviation of this thickness difference was calculated for each femur, along with the absolute mean.
Figure 7.10: Average cortical surface meshes. The posterior and anterior views of the mean outer and inner cortical surface meshes are shown in (a) and (b), respectively. Average closest distance between ground truth surface and mesh are shown as heat maps over the surfaces. The average difference between ground-truth thickness and mesh thickness is shown in (c). (d) shows mean mesh thickness of the validation set, mapped over the mean outer cortical surface mesh.
Table 7.1: Mean error statistics from the validation experiments.

<table>
<thead>
<tr>
<th>Meshes</th>
<th>Errors (mm)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMS</td>
<td>S.D.</td>
<td>Abs.</td>
</tr>
<tr>
<td>Leave-one-out</td>
<td>1.02</td>
<td>0.74</td>
<td>-</td>
</tr>
<tr>
<td>Outer Mesh (Ω_{out})</td>
<td>0.74</td>
<td>0.53</td>
<td>-</td>
</tr>
<tr>
<td>Inner Mesh (Ω_{in})</td>
<td>0.89</td>
<td>0.67</td>
<td>-</td>
</tr>
<tr>
<td>Mesh Thickness</td>
<td>0.61</td>
<td>0.60</td>
<td>0.03</td>
</tr>
</tbody>
</table>

The 17 scans were processed on a laptop computer with a quad core 2.4 GHz processor. Full processing time for each image was about 20 minutes. ASM segmentation failed to converge for two scans, so they were excluded from further analysis. One failure was due to poor initial positioning along the femoral shaft axis, while the other femur may have possessed an abnormally shaped proximal femur, since its shaft and distal regions were well segmented. Table 7.1 summarises the errors for the rest of the scans.

The outer cortical surface mesh (Ω_{out}) was, on average, less than 0.8 mm RMS from the corresponding Stradwin surface (Ω_{out}'). The accuracy of Ω_{in} was slightly worse than that of Ω_{out}. This was due to Ω_{in}'s dependency on Ω_{out} and the thickness field Ψ, so that error in Ω_{out} and Ψ were compounded. Despite this, the average thickness between Ω_{in} and Ω_{out} differed from Stradwin estimates by only 0.05 mm, with a standard deviation only marginally higher than that of Stradwin (0.58 mm). This indicates that cortical thickness modelled by the meshes preserved the accuracy of CTM well.

Figures 7.10a and 7.10b show mean errors over the average outer and inner cortical surface meshes, respectively. Error was below voxel resolution over most of the surfaces. The highest error was found on the femoral head. This was due to two factors. Firstly, the femoral head possessed very thin cortical bone, and in some cases, was undetectable in the images. Secondly, the nearby pelvis appeared much brighter in the images due to its higher density. In some cases, these two factors resulted in CTM mistaking the pelvic surface for the femoral head surface. The same problems were encountered in the condyles, with thin femoral cortex close to the patella and tibia. Despite these issues, error only exceeded voxel resolution in isolated regions.
Figure 7.11: Stradwin ground-truth limitations. The left panel shows the cortical thickness map of a validation femur calculated by Stradwin. Noisy estimates are apparent along the linea-aspera, and the femoral head. In the femoral head (upper panels), the extremely thin cortex causes Stradwin to mistakenly segment the cortical surface of the pelvis. In the femoral shaft (lower panels), Stradwin produced erroneous inner cortical surface due to surface normal mismatch. Smoothing constraints during mesh fitting minimised these error in our mesh.

Figure 7.10c shows average absolute error in mesh thickness. Errors were below measurement accuracy over most of the surface, even over the femoral head and condyle regions. Cortical thickness accuracy was thus not compromised by factors that lowered geometric accuracy. Thickness (and geometric) error was high on the linea aspera of the femoral shaft. However, these values are dubious due to unreliable CTM by both Stradwin and the presented method. This was caused by the sharp, ridge-like morphology of the linea aspera. A profile sampled normal to the outer surface may pass tangentially through the cortex, leading to an unreliable estimate of the inner cortical surface, and thickness. Smoothing penalties applied during inner mesh fitting mitigated this problem to a degree, giving the thick average linea aspera as shown in figure 7.10d as expected. Figure 7.11 illustrates this, as well as the issues in the
femoral head described above. Better ground-truth data, and a better definition of cortical thickness where external ridges are present, will be needed to improve the validation in these two regions.

Figures 7.12 and 7.13 show cross-sections through three meshes at various transverse slices. In both figures, from top to bottom, the meshes are of low, average, and high error, respectively. Faint thin cortex, and close neighbouring bones of the femoral head and condyles can be seen in these images. In the first column of figure 7.13, the linea aspera can be clearly seen, jutting away from the ellipsoid cross-section of the femoral shaft.

Figures 7.12b, 7.12d, and 7.12f show slices just above the lesser trochanter. In the upper right-hand portions of these images, cortical bone appears to reach outside of the segmented cortical surfaces. This apparent error can be due to a number of factors. Firstly, in this region, the cortex runs obliquely to the image slice plane as the femur expands from the femoral shaft into the femoral neck. The oblique angle makes the cortex appear thicker than its true thickness normal to the surface, which is how it is measured by CTM. Secondly, denser arrangement of cancellous bone is seen in this region, which at the resolution of the images, appears as if cancellous bone blends gradually into cortical bone. The lack of a sharp cancellous-cortical boundary means that manual interpretation will be highly subjective. The lack of the sharp boundary also violates the assumption of the CTM model (see Fig. 7.6). However, CTM does provide an objective measure, that as shown in the figures, is consistent across different femurs and not highly sensitive to dense cancellous bone near the cortex. Fitting of the CTM model is most sensitive to the initial positions of the inner and outer cortical surfaces. Fitting is more likely to be more successful if the initial positions are wider than the actual positions.

The high error case is of an 87 year old female, displaying signs of osteoporosis and osteoarthritis. The combination of abnormal morphologies produced noisy cortical surface location estimates. This is the cause of mesh creasing, as seen in figures 7.12e and 7.13f. Higher smoothing weights during fitting can reduce creasing. However, this comes at a cost of fitting accuracy in regions with high curvature.
Figure 7.12: Mesh outlines in selected image slices from the validation images. The left column shows slices through the femoral head. The right shows slices through the proximal metaphysis. The first row are from a segmentation with low error. The second row shows an average case, and third row shows a poorly-segmented case. Red and green dots are sampled from the outer and inner cortical meshes, respectively.
Figure 7.13: Mesh outlines in selected image slices from the validation images. The left column shows slices through the femoral shaft. The right shows slices through the condyles. The first row are from a segmentation with low error. The second row shows an average case, and third row shows a poorly-segmented case. Red and green dots are sampled from the outer and inner cortical meshes, respectively.
7.4 Full VIFM Set Segmentation

Cortical surface segmentation and meshing was performed on 262 normal CT images in the VIFM dataset to test reliability on a larger population. As stated in section 5.5.1, this set does not include images showing the presence of disease, prosthesis, or severe image artefacts.

No ground-truth data was available (except those in the training and validation sets), so a quality rating was given to each scans’ meshes, based on visual inspection. The ratings were:

A: no apparent defects.
B: apparent defects at 1 or 2 isolated locations.
C: apparent defects at 3 or more locations.
D: segmentation failed - femoral shaft not found, or ASM failed to converge.

7.4.1 Results

Out of the 262 scans process, 65 showed good quality meshes with no apparent inaccuracies or defects (A), 152 showed minor inaccuracies (B), 24 showed major inaccuracies or mesh defects (C), and 21 were not successfully segmented (F). In summary, 83% percent of scans were segmented and meshed with acceptable quality (A or B rating), and ASM segmentation converged in 92% of cases even if subsequent CTM produced poor results. Examples of meshes rated A, B, and C are shown in figures 7.14, 7.15 and 7.16 respectively.

When present, mesh defects were mostly found on the inner cortical surface mesh, along the linea aspera. As explained in section 7.3.2, thick ridges on the outer cortical surface decreased the reliability of CTM in finding the inner surface. The distal femur had the least number of defects. This may be due to the simpler geometry, and less variability compared to the proximal femur.

It should be noted that, despite removing 65 scans with the most severe abnormalities, significant variations still existed in the dataset. The dataset contained both male and female subjects, with an age range of 15 years to 95 years. Bones with various degrees of age-related bone loss and osteoporosis were present, as were wide variations
7.4. VIFM SCANS

Figure 7.14: Examples of A-rated cortical surface meshes segmented from the VIFM data set. These meshes had no apparent defects. The outer mesh (red) and the inner mesh (green) are superimposed on a coronal plane maximum-intensity projection of their respective CT image.

in cortical thickness. Leg posture was not strictly controlled in the VIFM images, so femur orientation was also highly variable.
Figure 7.15: Examples of B-rated cortical surface meshes segmented from the VIFM data set. These meshes had minor apparent defects. The outer mesh (red) and the inner mesh (green) are superimposed on a coronal plane maximum-intensity projection of their respective CT image.
Figure 7.16: Examples of C-rated cortical surface meshes segmented from the VIFM data set. These meshes had major apparent defects. The outer mesh (red) and the inner mesh (green) are superimposed on a coronal plane maximum-intensity projection of their respective CT image.
7.4.2 Enlarging ASM Training Set

To test whether the 41-femur training set for the ASM is sufficient, A rated femur meshes automatically produced by the pipeline were added to the training set, bringing the set size to 89 (7 A rated scans were already in the training set). The 262 scans were then re-segmented and meshed. There were no significant improvements to the success rate, or mesh quality, indicating that the 41-femur training set is sufficient.

7.5 Discussions and Future Work

In this chapter, we have presented a workflow for automatically segmenting and meshing the femur cortical surfaces, as well as mapping cortical bone thickness. The workflow enables two crucial aspects of femur morphology, shape and cortical thickness, to be automatically collected from CT images. Throughout the segmentation and meshing process, an average femur mesh is progressively fitted to segmented surfaces. Large deformations are constrained by a statistical shape model, which creates correspondent meshes across different femurs. This enables the meshes to be used for statistical analysis.

7.5.1 Training Set Size

An active shape model, trained with 41 femurs, was used to segment the femoral surface. According to our validation experiment, and the full VIFM dataset run, this was sufficient to segment over 90% of scans. No significant improvement in success rate was found by increasing the training set.

However, ASMs specific to certain populations may offer an improvement. For example, there is significant shape and appearance differences between sexes and age groups. ASMs trained with data specific to a particular group could be more sensitive and robust for scans from that group.

Lastly, mirroring the ASM and femur mesh would enable the right femur to also be meshed to double the training set size. This was not done since the training set was deemed sufficient, but will be straightforward to implement in the future. However, contra-lateral femurs must be treated with caution during shape model training.
7.5. DISCUSSION

Hislop-Jambrich (2010) showed evidence of subtle but significant shape differences between the left and right femurs. With such correlations, adding both to a training set would result in non-independent samples.

7.5.2 Suitability of ASM

The ASM method proved adequate for the presented method. In the leave-one-out experiment, ASM segmentations already deliver sub-pixel resolution accuracy. Recent developments, such as random forest regression (Lindner et al., 2012; Cuingnet et al., 2012), report higher accuracy than ASM. However, they are still reliant on the accuracy of the training data, and a CTM step would still be required for cortical thickness estimation. That said, is it feasible that ASM can be replaced by another robust segmentation method. Also, a statistical model could be used to constraint CTM parameters to improve its reliability, especially in regions with challenging morphology (e.g. femoral head).

7.5.3 Consequences of the Mesh

The piecewise smooth nature of the mesh has been exploited to filter erroneous data points. However, in regions prone to severe noise, such as along the linea aspera, or parts of the femoral head, smoothing was not always reliable in producing a fair mesh. Also, smoothing comes at a cost to the accuracy of the mesh in representing high curvature features. Figures 7.13a and 7.13c show mesh outlines around a section of the femoral shaft. It can be seen that the outer mesh rounds off the sharp corners of the linea aspera, and in figure 7.13d, there is higher than average error around the corners and grooves of the condyles.

Decreasing the weights on Sobolev and element boundary smoothing penalties during fitting will give more accuracy for these features. However, that would likely introduce more error in regions with noisy data. Alternative methods for reducing noise and improving mesh smoothness should be investigated in the future. That said, the required level of accuracy is application-specific. If statistical analysis of general morphology is the goal, then accurately representing small features will be less im-
important than robustly capturing large features, from which common morphometric measurements are made.

### 7.5.4 Computational Cost

Currently, the full segmentation process takes about 20 minutes on a quad-core 2.4 GHz laptop computer. There are a number of ways to improve computational speed. The majority of computational time is spent in conventional mesh fitting. ASM segmentation and CTM comprise about only a third of the full run time. Non-linear smoothing constraints mean that the fitting optimisation must be solved numerically. Also, a gradient function of the objective has not been derived. Linearising the smoothing constraints would reduce the fit to a linear problem, which would greatly improve run time. The segmentation and meshing pipeline is currently implemented in Python, with optimisations performed in C libraries through the SciPy package [Eric Jones et al.](https://www.scipy.org/). Moving completely to C or C++ could also offer some improvements in performance. Lastly, because each scan is segmented independently, many scans can be segmented in parallel. Implementing the pipeline on a cluster system would offer the greatest speed-up as long as segmenting a large population is an offline process.

### 7.6 Conclusions

This chapter has presented a fully automated system for segmenting and meshing femur cortical surface geometry, and measuring cortical thickness from clinical CT images. Given a template piecewise-parametric mesh of the femoral surface, active shape modelling and cortical thickness mapping are adapted to customize the mesh to inner and outer cortical surface geometries.

In validation experiments, the method produced meshes with an accuracy better than 0.9 mm RMS when compared to an established semi-automated technique. The method was used to process 262 images from the VIFM dataset, in which the outer cortical surface was segmented in 92% of the scans, and good quality meshes were produced from 83% of the scans.

Future work will focus on improving accuracy in small parts of the femoral head
and shaft. Also, we have proposed ways in which the performance of the system can be improved. In the next stage of the femur CT-to-model pipeline, the meshes produced are used to segment the cancellous bone regions, which is the focus of the next chapter.
Appendix 7
7.A Shaft Segmentation

The algorithm for automatically locating and segmenting the femoral shaft is detailed below. The output data points are used to initialise ASM segmentation (Sec. 7.2.1).

1. **Partition**: partition CT scan into overlapping 80 voxel\(^3\) cubes

2. **Filter non-bone blocks**: discard blocks with maximum intensity less than 50 HU

3. **Filter non-cylindrical structures**: for each remaining block
   a) Calculate centre of mass of each transverse slice
   b) Fit straight line to centres of mass of each slice
   c) Discard block if fit error > 1 mm R.M.S.

4. **Segment shaft**: for each remaining block
   a) Start with middle transverse slice
   b) Sample slice radially from its centre of mass, find shaft surface at the most negative gradient along the sample
   c) Fit low-degree spline to shaft surface points
   d) Propagate segmentation and spline fitting to slices above and below (can propagate beyond block)
   e) Continue until spline fit error rises above a critical value
   f) Discard blocks that have been segmented

5. **Point clouds**: group segmented points from adjacent slices into point clouds. Discard clouds with less than 1000 points

6. **Shaft point clouds**: combine remaining point clouds to form the shaft surface point cloud \(D^{\text{shaft}}\)
Chapter 8

Automatic Cancellous Image Registration

The final part of the femur CT-to-model pipeline segments and registers cancellous bone image volumes to a common reference space so that variation between femurs can be compared and quantified independent of shape. Cancellous bone is the major morphological component internal to the femoral cortex (Fig. 8.1), and plays an important role in femoral strength [Shim et al., 2007; Bryan et al., 2009; Langton et al., 2009]. Due to its porous structure, and therefore large surface area, the structure and distribution of cancellous bone are strongly affected by bone loss through ageing and osteoporosis. The effects are measurable in clinical quantitative CT images as changes in CT value, corresponding to volumetric Bone Mineral Density (BMD). Capturing these effects are crucial for CT-based tracking or diagnosis of osteoporosis [Fritscher et al., 2009; Laib and Regsegger, 1999; Holzer et al., 2009; Pitto et al., 2010], identification of fracture-prone femurs [Whitmarsh et al., 2011; Pulkkinen et al., 2004; Li et al., 2009; Bryan et al., 2009; Yang et al., 2012], or simply studying the differences in cancellous BMD between populations in general.
CHAPTER 8. CANCELLOUS REGISTRATION

Figure 8.1: Coronal slice through the proximal femur, with the cancellous bone region highlighted in red (a). However, much of the structural detail is lost in clinical-resolution CT images (b). Automatically capturing and comparing 3-D bone mineral density distribution is the focus of this chapter.

The registration process is illustrated in figure 8.2 and is composed of two steps:

1. Global non-rigid registration and segmentation based on inner cortical surface mesh correspondence, using Radial Basis Functions (RBFs);
2. Local non-rigid registration using the Free-Form Deformation (FFD) method of Schnabel et al. (2001).

Popular image registration methods such as Free-Form Deformation (FFD) Schnabel et al. (2001), or active appearance modelling Cootes et al. (1998) require manual, rigid-body, or affine initialisation, followed by computationally expensive iterative optimisation. In step one of the present method, RBFs are used to efficiently map cancellous bone images of arbitrary shape, position, and orientation to a reference space, by exploiting the correspondence of inner cortical surface meshes produced in the previous step of the pipeline (Cha. 7). The mapping is accurate enough so that in step two, only small-scale registration, using established methods, is needed to optimise correspondent between images. This differs from previous RBF image registration methods, which have used compact RBFs for local registration Siddiqui et al. 2009, Fornellet al. 1999, Wachowiak et al. 2004.
Figure 8.2: Overview of the cancellous bone registration process. Going from left to right, the cancellous bone volume of a CT image is firstly mapped to a reference space, using radial basis functions to interpolate the mapping between the reference and data meshes of the inner cortical surface. Then secondly, internal features of the image are registered against a reference image, using conventional free-form deformation.

Compared to FFD alone, the present method is significantly more accurate at the inner cortical surface (0.75 mm versus 3.7 mm), while requiring about a quarter of the run time (390 seconds versus 1600 seconds). Image correspondence was similar, if not better, than FFD alone, as measured by normalised mutual information and qualitative comparisons.

The next section will offer a brief recap of RBF theory, before the global RBF registration method is presented in section 8.2. The FFD registration step is briefly covered in section 8.3 before quantitative and qualitative validation results are presented and discussed in section 8.4. Registration was performed on 217 VIFM images to produce a preliminary PCA model of cancellous BMD distribution. This model is presented and discussed in section 8.5. Finally, conclusions and future work are discussed in section 8.6.
8.1 Radial Basis Functions

The value of an RBF field $R$ at point $d$ is given by

$$R(d) = \sum_{i=1}^{Q} a_i \phi(\|d - c_i\|)$$ (8.1)

where $c_1, \ldots, c_Q$ are the coordinates of RBF knots, $a_1, \ldots, a_Q$ are their respectively scalar coefficients, and $\phi$ is an RBF whose value only varies with respect to the radius $\|d - c_i\|$. For the methods below, Gaussian RBFs are used:

$$\phi(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right)$$ (8.2)

Following the method of Benoudjit et al. (2002), the Gaussian width $\sigma$ for each knot is calculated as the average distance to its nearest three knots, scaled by a factor $s$.

The RBF coefficients are calculated by fitting the RBF field to data points with values $u_1, \ldots, u_D$ and coordinates $d_1, \ldots, d_D$ by solving the linear system

$$
\begin{pmatrix}
\phi_{1,1} & \cdots & \phi_{1,Q} \\
\vdots & \ddots & \vdots \\
\phi_{D,1} & \cdots & \phi_{D,Q}
\end{pmatrix}
\begin{pmatrix}
a_1 \\
\vdots \\
a_Q
\end{pmatrix}
=
\begin{pmatrix}
u_1 \\
\vdots \\
u_D
\end{pmatrix}
$$ (8.3)

where

$$\phi_{i,j} = \phi(\|d_i - c_j\|)$$ (8.4)

For more details about RBFs, please see section 4.6.

8.2 Global RBF Registration

The global registration involves an RBF field providing a non-rigid mapping between the cancellous bone volume in a data image, and a reference space (Fig. 8.3). The reference space is bound by the mean inner cortical surface mesh (reference mesh), while the data image space is bound by the data inner cortical surface mesh (data mesh). The correspondence of the meshes means that a mapping already exists between the two spaces on their surfaces. A Gaussian RBF field $R$ is used to interpolate this mapping across the internal volume such that a point $d$ in the reference space is mapped to $d'$ in the data space by

$$d' = R(d)$$ (8.5)
Figure 8.3: Cancellous bone volume registration and segmentation using radial basis function mapping. Voxel coordinates in the reference space are mapped to the data space by the RBF field $R$ to get CT values from the data image. The knots of $R$ (red spheres) are evenly distributed in and around the reference cancellous bone volume, bound by the reference inner cortical surface mesh. Knot coefficients are fitted to data points on the reference mesh (green spheres), who’s values are their corresponding coordinates on the data mesh.

The mapping is completely described by the RBF field, and no prior rigid or affine alignment of the reference and data spaces is required.

$R$ contains 159 knots, spaced 15 mm apart within and around the reference cancellous bone volume, which is bound by the reference mesh. For each femur, $R$ is fitted to a dense distribution of data points $d^\text{surf}$ on the reference mesh (2.5 mm spacing, 3583 points), whose values are their corresponding coordinates $d^\text{surf}'$ on the data mesh. The same set of data points is used for every femur. The way in which knot spacing and RBF width scaling were determined is presented below in subsection 8.2.1.

Once RBF weights have been fitted, the BMD value at $d'$ in the data image $I'$ is sampled and mapped to its corresponding positions $d$ in the reference image $I^\text{RBF}$:

$$I^\text{RBF}(d) = I'(R(d))$$

$I^\text{RBF}$ is reconstructed at 1 mm resolution.

The reconstructed image excludes voxels within 1.5 mm the inner cortical surface to remove the presence of any cortical bone at the boundary of the cancellous volume.
Exclusion depths of 1.0 mm, 1.5 mm, and 2.0 mm were tried, and 1.5 mm was found to be sufficient. Cortical bone appears due to errors in the inner cortical surface mesh, and leads to spurious patches of high BMD at random locations on the cancellous volume. These patches will adversely affect the subsequent FFD registration and skew statistical models of the cancellous BMD distribution.

### 8.2.1 Determining Knot Spacing and Width

Optimum knot spacing of 15 mm, with RBF width scaling $s = 60$ was determined by fitting $R$ on 30 randomly selected femurs, using a range of knot spacings and RBF width scaling values. Denser knots may improve $R$ fitting accuracy, and therefore mapping accuracy at the surface, but decrease the stability of the mapping across the internal volume. As shown in figure 8.4, this causes a distorted mapping where points in the reference volume are erroneously mapped, some ending up on the outside of the data volume.

![Figure 8.4](image)

**Figure 8.4:** Effects of knot spacing and width on mapping a regular grid of points. (a) Small knot spacing and RBF width leads to a distorted mapping. (b) Higher knot spacing and RBF width preserves the regular topology.

For each femur, for each knot spacing-width combination, two errors were calcu-
8.2. GLOBAL RBF REGISTRATION

lated:

**Surface Error:** $R$ fitting error, i.e. the RMS distance between points mapped from the reference mesh to the data space $R(d_{\text{surf}})$, and the corresponding points on the data mesh $d_{\text{surf}}'$:

$$
\epsilon_{\text{surface}} = \sqrt{\frac{\sum_{i=1}^{D} \| d_{\text{surf}}'_{i} - R(d_{\text{surf}}_{i}) \|^2}{D}}
$$

(8.7)

**Volume Error:** the sum of the closest distances to the data mesh of points mapped outside the data mesh. A regular grid of points $d_{\text{vol}}$ in the reference volume was mapped to the data volume $R(d_{\text{vol}})$. For mapped points lying outside of the data mesh, the volume error was the sum of their closest distances to the mesh.

![RBF Mapping Surface Error](image)

(a)

![RBF Mapping Volume Error](image)

(b)

**Figure 8.5:** Average surface (a) and volume (b) errors for different knot spacing and width scalings values.
The results from these tests showed that surface error was minimal when RBF width scalings $s$ was between 10 and 20, and gradually increased with $s$ (Fig. 8.5a). Also, as expected, surface error decreased with knot spacing: more knots closer together provided more degrees of freedom to minimise errors. This held true until too many degrees of freedom caused the fitting problem to become under-constrained and unstable.

Volume error displayed a sharp drop-off around a critical value of $s$ for each knot spacing, with smaller spacing requiring a larger $s$ (Fig. 8.5b). Above this critical $s$ value, volume error was relatively invariant between the three finer knot spacings.

The results suggest that there is a knot spacing-dependent minimum $s$ for stable volume mapping. This is likely due to the role of $s$ in determining the RBF width of each knot, which was calculated by multiplying $s$ with the average distance to the knot’s three nearest neighbouring knots. As this distance decreases with knot spacing, $s$ must increase to maintain the same RBF width to interpolate the mapping across the cancellous volume. Since surface error increases with $s$, there is a necessary trade-off in surface accuracy for volume mapping accuracy.

Knot spacing of 15 mm (giving 159 knots), with $s = 60$ gave a good balance between surface error and volume error. Knot spacings of less than 15 mm, while producing smaller errors at the volume surface, resulted in internal distortions in many images, which were unpredictable and hard to control. These distortions are illustrated in figure 8.4. It may be that the number of knots required for 12.5 mm and 10 mm spacing over-fitted the problem. With 15 mm spacing, width scaling of 60 produced the smallest volume error. Its average surface error of 0.76 mm was acceptable since it was below the image resolution of 1 mm $\times$ 1 mm $\times$ 1.6 mm, and within the margin of error of the inner cortical surface mesh (Sec. 7.3), especially when taking into account the 1.5 mm separation between the mesh and reference image sampling points.

8.3 Local FFD Registration

After global registration, a local registration step is required to fine-tune the correspondence of features in the cancellous bone volume. This is necessary because the
global registration only considers the correspondence of the volume surface. Following the method of Schnabel et al. (2001), and using the IRTK\(^1\) software, the globally registered image \(I^{RBF}\) is embedded in a B-spline (Catmull and Clark, 1978) lattice with nodes (control points) \(X\), which deforms the image to match a reference image \(I^{ref}\):

\[
I^{ref} \approx I^{FFD} = T^{FFD}(X, I^{RBF})
\] (8.8)

where \(T^{FFD}\) minimises

\[
\epsilon^{FFD} = -\Upsilon(I^{FFD}, I^{ref})
+ \lambda_{\text{smooth}}\epsilon_{\text{smooth}}(X)
+ \lambda_{\text{volume}}\epsilon_{\text{volume}}(X)
+ \lambda_{\text{topology}}\epsilon_{\text{topology}}(X)
\] (8.9)

The similarity between \(I^{FFD}\) and \(I^{ref}\) is measured by \(\Upsilon\) - their Normalised Mutual Information (NMI) (Studholme et al., 1999). Briefly, maximising NMI maximises the overlap in information (or features) between the two images, relative to the information present in each. NMI allows image similarity to be quantified without the need for landmarks, which are difficult to find in highly variable cancellous bone.

Going back to equation 8.9, the remaining three terms are weighted penalty terms preserving B-spline lattice smoothness, volume, and topology.

FFD local registration is performed at three resolutions: 8 mm, 4 mm, and 2 mm, with B-spline lattice spacing of 80 mm, 40 mm, and 20 mm, respectively, to allow efficient registration of coarse to fine features. Since the cancellous bone volume in both \(I^{RBF}\) and \(I^{ref}\) already fully overlap in terms of volume (not in terms of internal features), the B-spline lattice external to the volume is fixed (Fig. 8.6) and a high weighting is applied to the volume-preservation penalty \(\epsilon_{\text{volume}}\) in equation 8.9. The reference image \(I^{ref}\) is the globally registered image of a representative femur.

8.4 Validation

Thirty randomly selected femurs were registered to quantify the surface error, image similarity, and run time of three registration methods:

\(^1\)http://www.doc.ic.ac.uk/ dr/software/index.html
RBF: RBF global registration only;
RBF+FFD: RBF global registration followed by FFD local registration;
FFD: FFD registration with affine initialisation, which was the benchmark.

FFD registration with manual or automatic initialisation is commonly used for statistical analysis of a population of images, as demonstrated in Loeckx et al. (2003); Rueckert et al. (2003); Whitmarsh et al. (2010, 2011); Carballido-Gamio et al. (2011), making it a good benchmark to compare against.

RBF and RBF+FFD registrations were performed following the descriptions in sections 8.2 and 8.3. FFD with affine initialisation is described below.

### 8.4.1 Benchmark FFD Registration

FFD-only registration was carried out using IRTK, using an average case reference femur to provide a reference inner cortical surface mesh, and a reference image. This was the same reference femur as the one used for RBF+FFD registration. For each data femur:

1. The data image was cropped to a rectangular volume surrounding the data inner cortical surface mesh, padded by 10 pixels per side.
2. The affine transformation registering the data mesh to the reference mesh was applied to the cropped image as its initial transformation.

3. Affine registration was refined using IRTK at 12 mm, 6 mm, and 3 mm resolution.

4. IRTK FFD registration was performed using the same parameters as for RBF+FFD registration (8 mm, 4 mm, and 2 mm resolution, with B-spline lattice spacing of 80 mm, 40 mm, and 20 mm respectively).

5. The registered data image was sampled at a grid of points within the reference mesh, with a 1.5 mm gap to the reference mesh, to produce a cancellous bone volume image.

8.4.2 Comparison Metrics

Surface Error For RBF and RBF+FFD registration, registration error at the cancellous volume surface was simply the RBF fitting error from equation [8.7]. For FFD registration, the error was the RMS distance between points sampled on the reference mesh, and corresponding points sampled on the data mesh transformed by $T_{FFD}$. The points were the same set of 3583 evenly distributed points used for fitting the RBF registration.

Image Similarity To measure the similarity of images registered using each method, NMI was calculated between each pair of registered images, giving 435 measurements per method.

Run Time The run time of each image was recorded for each method. The time includes registration and image sampling time. Registrations were all performed on a 2.4 GHz quad core desktop computer with 4 GB of RAM.

8.4.3 Results and Discussion

As shown in figure [8.7], RBF and RBF+FFD registration had significantly lower surface error and run time compared to FFD registration. RBF registration produced lower image similarity than FFD, but FFD after RBF registration (RBF+FFD) improved image similarity beyond the level of FFD-only registration.
CHAPTER 8. CANCELLOUS REGISTRATION

Figure 8.7: Comparison of surface error (a), pairwise normalised mutual information (b), and run time (c) between RBF, RBF+FFD, and FFD-only registration.

Mean RMS surface error was 0.76 mm for RBF and RBF+FDD, compared to 3.7 mm for FFD. This can be attributed to RBF registration where error between well-defined surfaces (meshes) was explicitly minimised while fitting the RBF. FFD-only registration lacked such a constraint (except for the initial affine transformation), and was purely dependent on matching image features that could vary in appearance and shape.

Mean run time was 10.1 seconds for RBF, 390 seconds for RBF+FFD, and 1583 seconds for FFD registration. Compared to similar work in literature, the RBF+FFD method showed superior performance. In the original FFD registration work [Rueckert et al. (1999)], breast MRI images were registered in 15 to 30 minutes on a workstation of that time. In more recent work, [Whitmarsh et al. (2011)] registered proximal femur images in approximately 15 minutes, and [Bryan et al. (2010)] registered whole femurs in approximately 30 minutes, both on modern desktop computers. The reduction in run time of RBF+FFD compared to FFD was due to two reasons:

1. RBF registration brought the data and reference images into close alignment, so that fewer iterations of subsequent FFD registration were required;
2. In RBF+FFD registration, the image external to the accurately aligned cancellous volume surfaces didn’t need to be considered in the FFD stage (because
of accurate RBF alignment at the surface). This meant that the number of B-spline control points used in RBF+FFD was greatly reduced (see Fig. 8.6), which reduced the degrees of freedom in the optimisation problem. FFD registration alone showed poor results unless external features (cortical bone, soft tissues) were included in the image.

Low surface error and run time indicates that RBF-global registration is an efficient method to ensure image surface volume registration, which vastly reduces the amount of computation required for internal image registration.

As expected, image similarity (in terms of NMI) after RBF registration was lower than that of FFD registration, since internal image features had not been specifically registered in the former. Subsequent FFD in RBF+FFD was able to provide a slightly better registration than FFD alone. Since NMI was used in the registration itself (by FFD), it could be seen as a biased measure. However, there was great difficulty in finding other suitable metrics. One measure would be the error between manually defined internal landmarks in the target and registered images. However, due to the variability of cancellous structures, and the smooth spatial variation in BMD, only a small number of landmarks would be identified consistently. The fine structure of cancellous bone makes measures similar to cross-correlation overly sensitive to small misalignments, and commonly used overlap percentage measures do not measure the correspondence of internal features. Further work is required for more objective measurements of internal correspondence.

A qualitative comparison is presented in the composite coronal views of the proximal (Fig. 8.8) and distal (Fig. 8.9) femur. In the proximal slices, each square sub-image is taken from one of the registered images. Compared to the FFD composite, the RBF and RBF-FFD composites show better alignment of anatomical features across adjacent sub-images, especially in the femoral head and the volume boundary. As indicated by the red arrows in figure 8.8d, cortical bone was mis-registered into the cancellous volume in some cases using FFD-only registration, which was not seen in the RBF and RBF+FFD images.

In the distal composite images (Fig. 8.9), each column (separated by green dotted lines) is taken from a different registered image. The yellow dotted line indicates the
Figure 8.8: Composite slices of registered proximal femur cancellous volume. (a) is the reference image, while (b), (c), and (d) are composite slices made of some of the 30 images registered using RBF, RBF+FFD, and FFD respectively. Each square sub-image in the composites is taken from one of the registered images, all at the same coronal slice. The epiphyseal lines in the femoral head and greater trochanter are labelled with yellow dotted lines where identifiable. Image intensity varied greatly between images, but the RBF and RBF+FFD composites show better alignment of anatomical features across adjacent sub-images. Red arrows in the FFD image indicate the presence of cortical bone in the cancellous volume, suggesting poor registration.
Figure 8.9: Composite slices of registered distal femur cancellous volume. (a) is the reference image, while (b), (c), and (d) are composite slices made of 8 of the 30 images registered using RBF, RBF+FFD, and FFD respectively. Each column sub-image (separated by green dotted lines) in the composites is taken from a registered image, all at the same coronal slice. The distal epiphyseal line is labelled by a yellow dotted lines where identifiable. RBF and RBF+FFD composites show better alignment of anatomical features across adjacent sub-images, especially the epiphyseal line in the RBF+FFD image. Red arrows in the FFD image indicate the presence of cortical bone in the cancellous volume, suggesting poor registration.
distal epiphyseal plate. This plate was better aligned across different images registered using RBF+FFD than using FFD alone. Also, the general appearance of the RBF and RBF+FFD composites better matched the reference image than the FFD composite. Again, red arrows indicate cortical bone in FFD images where they are not seen in the reference, RBF and RBF+FFD images.

8.5 Full VIFM Set

The cancellous image volumes of 217 femurs were registered using the RBF+FFD method above. These femurs had well-segmented inner cortical surface meshes, as validated in the previous chapter (Sec. 7.4). As a further step of qualitative validation, the mean and principal components of the registered images were calculated, and presented below. Well-registered images should produce a mean image in which major anatomic features are distinguishable. Principal components of the image should reflect distinct modes of variations, and their variations should exhibit the typical J-shaped curve, as opposed to a flat curve exhibited by random data.

Principal Component Analysis (PCA) (Sec. 4.1) on the registered images was carried out as follows:

1. The mean image was calculated and subtracted from each image;
2. For each image, cancellous volume voxel values were placed in a column vector (background voxels were ignored). Columns for all images formed the PCA data matrix;
3. The first 100 principal components were calculated using the Non-Linear Partial Least Squares (NIPALS) algorithm (Wold 1975);
4. Horn’s parallel analysis (Horn 1965) was performed to determine the number of useful principal components. This involved plotting principal component variance versus that of randomised data (Fig. 8.10).

Real data variance intersected random data variance at the 20th component, which covered 64% of total variation. This suggests 20 useful components for cancellous BMD reconstruction and analysis. Figure 8.11 shows the mean and variations in the first five principal components, in the proximal and distal femur. Anatomical structures such as
epiphyseal plates are still preserved in the mean images, as are spatial variations similar to those in the reference image in figures 8.8a and 8.9a. Each of the five components are briefly described below:

- Component one accounted for a global change in BMD value, accentuated where average values are already higher (e.g. femoral head, neck, and vertical “bundles” in the condyles);
- Component two accounted for a milder global change, with minimal or no change at epiphyseal plates;
- Component three coupled an increase in BMD in the superior femoral neck and lower (distal) proximal femur, with a decrease in the femoral head, greater trochanter, and the distal femur;
- Component four showed BMD increase in most regions, except in the joint-loaded regions of the head and condyle, where there was no change, or a slight relative decrease;
Component five showed BMD changes in the epiphysis (head, greater trochanter, condyles) opposed by changes in the femoral neck and the upper (proximal) distal femur.

It should be noted that the directions of change are invertible within each component, and it is the relative change between regions that is of interest.

Due to time constraints, these variations have not been correlated to factors such as age and gender, which could be responsible for the variations represented by the first two principal components. It is encouraging that the variations of components three, four, and five occur in anatomically interesting regions. Femoral neck BMD is well correlated to fracture risk [Whitmarsh et al., 2011; Pulkkinen et al., 2004; Danielson et al., 2012], while BMD near joint surface (subchondral BMD) is known to be affected by exercise in animals [Kawcak et al., 2000; Murray et al., 2006; Oettmeier et al., 1992] and osteoarthritis [Pastoureau et al., 1999; Hayami et al., 2004; Petersson et al., 1998]. The results suggests that the variation are clinically significant, but more patient information and pathological cases are required for verification.

The BMD PCA model is much less compact when compared to the PCA shape model (Sec. 6.3.5). However, this observation is in line with similar work in literature. In [Bryan et al., 2010], with a training set of 46 femurs (43 to 91 years of age), a PCA model of shape and BMD explained 95% of variance in 35 significant components. In [Whitmarsh et al., 2010], with a training set of 60 proximal femurs (mean age of 79 ± 11), a PCA model of BMD explained 60% of variance in 11 significant components. The 217 VIFM images is significantly larger and more variable than the training set in these similar works, which contributes its relatively lower compactness.

As suggested by [Bryan et al., 2010], the noisy nature of CT images and cancellous bone structure probably contributes greatly to the poor performance of PCA. Cancellous bone exhibits a highly variable struts and plates structure which is impossible to register perfectly between individuals. When voxel size is at the scale of these struts, it means that the intensity of each registered voxel varies hugely between bone and soft tissue values. Another factor reducing compactness is the large number of parameters used to describe BMD distribution. Whether using registered image voxels or tetrahedral mesh vertices, tens to hundreds of thousand of parameters are used. This
Figure 8.11: Mean and principal components of 217 cancellous bone images. Images shown are coronal slices of the proximal and distal femur. Each pair of images (labelled PC1 to PC5) shows 1 standard deviation of variation from the mean along its labelled component. Red indicates increase in CT value, and blue the opposite. The direction of change within each component is invertible, so it is the relative change between regions that is interesting. Non-femur pixels are dark to highlight the femur region, and do not actually vary along any component.
oversampling introduces noise into the BMD description and statistical analysis.

8.6 Conclusions and Future Work

This chapter concluded the femur CT-to-model pipeline, with methods for segmenting and registering the cancellous bone volume, which was bound by the inner cortical surface mesh produced from the previous step of the pipeline. Registration was composed of an RBF-based non-rigid global registration, followed by a conventional freeform deformation local registration. We have shown that the RBF global registration significantly improved the accuracy and run time of the registration compared to FFD alone. General anatomical structures were well registered across 217 VIFM femur images, and PCA revealed variations in anatomical regions relevant to fracture risk and osteoarthritis, which warrants further investigation.

Optimising RBF knot placement is a path of future work for the registration method. Currently, knots are placed in a uniform grid, which gives acceptable overall registration accuracy. An optimised knot distribution could improve accuracy in regions with high geometric variability, while reducing the number of knots needed (although computational cost is not an issue in the current implementation).

Another possibility for future work is to replace local FFD registration with an RBF-based local registration method, for example Fornefett et al. (1999); Wachowiak et al. (2004), or Siddiqui et al. (2009). This will simplify the registration process, and further reduce computational cost, since the B-spline lattice will be replaced by RBF knots placed only where needed.

Reducing the number of parameters for cancellous BMD distribution may improve the compactness of subsequent statistical models. Currently, each registered image contains over 270,000 voxels, or parameters for statistical analysis. Other than simply down-sampling the images, other approximation methods could be used to reduce this number. For example, smooth 3-D piecewise-parametric meshes have been used to interpolate BMD (Shim et al. 2007). Local structural measures such as trabecular orientations, spacing, or fabric tensor can also be interpolated to compensate for small intensity-encoded features lost by smooth BMD interpolation. Alternatively, radial
basis functions have also shown promise in image compression \cite{Magoulès2007, Vaneček2010}, and can be applied here for 3-D images.

The next chapter will discuss future work goals for correlating registered cancellous volumes with cortical surface meshes. This will combine models of femur shape with BMD distribution, so that complete femur morphologies can be generated for finite-element modelling, or regressed against patient information for classification and prediction.
Chapter 9

General Discussion, Conclusions, and Future Work

The principal aim of this thesis was to develop methods for segmenting, modelling, and statistically analysing femur morphology, from a large population of CT images. The pipeline aimed to utilise statistical shape modelling methods, and was targeted at a population of 262 CT images provided by the Victorian Institute of Forensic Medicine.

An automatic femur CT-to-model pipeline has been developed, which meets the above goals in the following aspects:

1. A statistical shape model of the femur has been created using a novel region-based meshing and training method;
2. This shape model was used to drive automatic meshing of the femoral cortex through a novel combination of active shape modelling and cortical thickness mapping;
3. Meshing of the inner cortical surface has enabled a novel method of registering cancellous Bone Mineral Density (BMD) image using radial basis functions.

By the above methodology, femur morphology in terms of cortical surface geometry, cortical thickness, and cancellous BMD distribution, is automatically extracted and modelled.

Future work will primarily involve integrating the three morphological models for (a) generating full femur models, and (b) statistical analysis to further the understand-
ing of femur morphology variation.

9.1 Key Achievements

9.1.1 Femur Shape Model

As detailed in chapter [1], a statistical shape model of the femur was produced in the training phase of the CT-to-model pipeline. Corresponding regions on a training set of 41 femoral surfaces were automatically partitioned and grouped using region-growing and mean-shift clustering. These regions were used to design a region-based quartic-Lagrange femur mesh, which was fitted region-by-region to manually segmented surfaces to train the femur statistical shape model.

This region-based shape model was fitted more accurately during training (0.52 mm RMS), and produced more correspondent reconstructions using less principal components in leave-one-out experiments, compared to an equivalent non-regional model.

The shape model above was crucial in the CT-to-model pipeline for automatically meshing the femoral cortex.

9.1.2 Automatic Cortex Meshing

Cortical bone geometry was automatically extracted and modelled in the first step of the processing phase of the pipeline (Cha. [2]). ASM and CTM were adapted for the quartic-Lagrange femur mesh, and combined to mesh the inner and outer cortical surfaces to sub-pixel accuracy.

Segmented meshes were accurate to 0.9 mm RMS, and cortical thickness to 0.6 mm RMS, compared to semi-automatic segmentation by the state-of-the-art software, Stradwin. Segmentation of 262 VIFM images produced 217 well-meshed femurs, a success-rate of 83%. Future work will aim to improve robustness, and reduce computational cost.
9.1.3 Automatic Cancellous Image Registration

Cancellous BMD image was automatically registered in the second step of the pipeline processing phase (Cha. 8). Images were mapped to a reference volume by radial basis functions (RBFs), which interpolated the mapping between segmented inner cortical surface meshes, and a reference inner cortical surface mesh. This global non-rigid registration was accurate enough so that only minor further adjustments using conventional Free-Form Deformation (FFD) was required.

RBF then FFD registration led to a four-fold reduction in run time compared to FFD-only registration, and a surface accuracy of 0.76 mm, compared to 3.7 mm for FFD-only. Registration of 217 images showed good alignment of variable anatomical structures, and the principal components of registered images showed BMD variations specific to anatomic regions, which hints at interesting relationships that warrant further study.

9.2 Limitations and Future Work

Limitations and possible improvement for each specific pipeline component have been discussed in their respective chapters. This section will discuss general limitations and possible future work.

9.2.1 Limitations of the Dataset

The presented CT-to-model pipeline was designed for and test only on 262 CT scans from the VIFM dataset. In addition, scans with abnormal femurs and artefacts were excluded from this set. The consequence was that

- The system is untested on CT scans with different imaging parameters, and other imaging modalities such as MRI;
- The system’s robustness to imaging artefacts is unknown;
- The system’s robustness to abnormal or pathological femur morphology is unknown, but possibly poor. The statistical models produced only captured variations present in normal morphologies. Abnormal morphologies may be of more
interest in clinical research.

The pipeline was tested on a set of scans which exhibited a wide range of variation in terms of subject age and size, femur position, orientation and morphology. The CT scans themselves are representative of common post-mortem CT scans, and the subjects are representative of an urban Western population (the State of Victoria, Australia). In addition, the 262 scans used are a small fraction of the tens of thousands of full-body scans stored at the VIFM. Thus, with the current pipeline, not only can tens of thousand of femur be extracted, but with retraining and remeshing, also thousands of other bones.

The pipeline can be easily modified to process a set of scan with different imaging properties such as image resolution or X-ray energy. Since the appearance of the femur will be different, the ASM will need to be retrained, and parameters of FFD image registration may need to be adjusted. If abnormal femurs or another bony structure is of interest, then the statistical shape model will also need to be retrained to capture different shape variations, and cortical thickness mapping parameters may need to be adjusted to accommodate different cortical thickness distribution. The overall framework of the pipeline should not require modification.

As mentioned above, it would be advisable to retrain the pipeline to cater to the specific variations found in the abnormal femurs. Adding abnormal morphologies to the current normal morphology shape model would be detrimental because this would create a highly non-Gaussian shape and appearance distribution in the training set. Since PCA is sensitive to outlier, the resulting principal components may poorly model both normal and abnormal populations. If the pipeline is to process a set of scans containing both normal and abnormal femurs, a classification step should be included to choose the appropriate pipeline parameters (such as shape and appearance models and CTM parameters) for the class of the incoming femur. Classification can be based on image metadata, or the results of segmentation using generic pipeline parameters. Besides normal and abnormal, classification can also include different sexes or age groups if more specific choices of parameters lead to more accurate or robust processing.
9.2. LIMITATIONS AND FUTURE WORK

9.2.2 Training Set Size and Statistical Methods

The pipeline used a PCA model of 31 femurs to drive its segmentation component. The size of the training set appeared sufficient for this purpose. As stated in section 7.4.2, segmentation results did not change when the training set was increased to 89. A more rigorous cross-validation experiment could be performed to determine the optimum training set size for the shape model. PCA was an appropriate statistical modelling method since ranked orthonormal principal components made for very efficient fitting during ASM segmentation, using only 5 to 6 components. Normal femur morphologies meant that assumptions of normality were applicable.

For the cancellous BMD appearance model, the optimal training set and statistical method is less certain. The PCA model of BMD distribution was not used for predictive purposes, so no approximation error values were available. The model (Sec. 8.5) explained 64% of variations in 20 significant components, which is comparable to similar work when accounting for its large and variable training set. The BMD statistical model is much less compact than the shape model, which can be attributed to a combination of noisy CT image characteristics, high number of appearance parameters (voxel values), and registration errors. As described in section 8.5, a possible solution is describing BMD distribution as smoothly varying fields of BMD and local cancellous structural information.

Choice of the method of dimension reduction depends on the downstream application and the data itself. Benefits of PCA for data reconstruction and approximation has been stated before. For prediction and classification, it would be necessary to compare PCA to other methods to determine the best method. ICA is known to more sensitive to local variations while PCA describes global variations (Suinesiaputra et al., 2009). This suggests that ICA may be more appropriate for identifying local changes in BMD that may be crucial in determining fracture risk. The linearity of BMD variation is unknown, especially with the effects of ageing, menopause, and osteoporosis. Testing BMD variation linearity would be a worthy investigation before applying non-linear methods such as kernel PCA. Alternatively, linear methods may still be applicable is the population is appropriately partitioned based on sex, age, and other factors.
Shape or appearance descriptions with too many parameters introduce noise into statistical models. In the current pipeline, no effort has been made to find the optimum number of parameters needed to describe either shape nor appearance. An objective way to determine optimality is to use information criteria methods, such as Akaike information criteria (Sakamoto et al., 1986), Bayesian information criteria (Chen and Chen, 2008), or minimum description length (Barron et al., 1998). These methods can be used to measure the goodness of fit of a description with a particular number of parameters. Assuming a mesh-based piecewise parametric field description of shape and appearance, information criteria can be used to control refinement of the mesh until optimum description is achieved.

9.2.3 Pipeline Success Rate and Accuracy

The CT-to-Model pipeline had an overall success rate of 83% on the 262 CT scans, which was heavily dependent on the robustness of cortex segmentation and meshing. Beginning with identifying the femoral shaft, errors in each step reduced the robustness of following steps (ASM segmentation, CTM, RBF registration, FFD registration). Improvement or replacement of earlier steps will have a greater effect on improving overall robustness. The Hough transform (Ballard [1981]) and random forest regression using Haar-like features (Cootes et al., 2012) are promising methods for improving the shaft-identification and segmentation steps.

Pipeline accuracy can be looked at independently for each component. For mesh fitting during training, sub-voxel error was acceptable since the manual segmentations are only accurate to 1 voxel. For cortex meshing, error was influenced by both CTM error and mesh fitting error. CTM error was around 0.6 mm RMS, while fitting error was around 0.5 mm RMS, therefore an error under \( \sqrt{0.6^2 + 0.5^2} = 0.78 \text{mm RMS} \) was acceptable. This was achieved by the pipeline for thickness measurements and the outer cortical mesh, and closely matched by the inner cortical mesh. For registration, error at the image boundary was expected to match the cortex meshing error (actual accuracy was 0.76 mm RMS). However, there is a bit more leeway due to the 1.5 mm buffer between the inner cortex mesh and the cancellous image. Internal registration error was hard to qualify especially without manually defined landmarks. The best measure
may be the quality of the BMD statistical model produced.

9.2.4 Statistical Relationships Between Morphology Models

Currently, the pipeline produces three models of femur morphology per image: cortical surface geometry, cortical thickness, and cancellous BMD distribution. Cortical geometry was separated from the cancellous BMD volume because surface correspondence was easier to achieve through surface landmark features. Volumetric correspondence was harder to achieve since internal volumetric features are much more variable. Therefore, the surface was decoupled to optimise its correspondence independent of the interior volume.

Three separate models allow statistical modelling of each morphological aspect independently, but an obvious extension is a statistical model relating all three aspects, using methods such as principal component analysis, partial least-square regression, or canonical correlation analysis. This will be useful in understanding the relationship between femur shape and bone mass distribution (in cortical and cancellous bone), as well as predicting and generating complete femur morphologies from partial information.

A more radical solution to combining the morphological models is a redesign of the morphology description. The surface mesh can be extended to a volumetric mesh to capture volumetric properties such as BMD and cancellous structure (see Sec. 8.6).

9.2.5 Traditional Morphometrics

Traditional measurements of femur lengths, widths, angles, and BMD at specific sites can be automatically taken on the cortical meshes and cancellous BMD images generated by the pipeline (Fig. 9.1). Such measurements provide familiar summaries of 3-D morphology to clinicians, and enable results obtained by the pipeline to be compared to older studies.

Material positions on the cortical meshes generally correspond to anatomic features, which are used for these measurements. Thus femoral lengths, width, and axes can be easily calculated from predefined material points on the meshes. Registration of cancellous BMD images to a reference space means that BMD at a specific anatomic site can be taken from the same reference space coordinates across all registered images.
Figure 9.1: Preliminary data from automatic morphometry measurements on the VIFM images. Femoral head diameter and femoral length distributions for males and females are as expected.
The software framework for defining and taking measurements in the way described above has already been implemented in the pipeline. However, the precision and accuracy of the measurements need to be validated in future experiments.

9.2.6 Correlation to Anthropological Data

As an extension of the goal above, another point of future work is to relate morphological variations to anthropological information, such as subject age, gender, weight, and lifestyle factors. This will enable the generation of femur morphologies representative of specific populations, which is useful for improving segmentation and registration convergence, and downstream analysis tasks such as finite-element modelling and prosthesis design. Another major area of application is in forensic sciences, where femur morphology is used for victim identification. Accurate automatic prediction of information such as age and gender from recovered femurs can reduce the labour on workers in disaster or mass-grave situations.

9.2.7 Application to Other Skeletal Structures

The general methods developed for this thesis are not specific to the femur - the shape modelling, segmentation, and registration methods can be readily applied to other bones. This generality enables the pipeline to extract the morphologies of other bones.

The hope is to make automatic extraction and modelling from large datasets standard and routine. This way, detailed morphology can be readily studied not just on individual bones, but across different bones and other organs.
References


REFERENCES


BodyParts3D. Anatomography, 6 January 2013.


REFERENCES


REFERENCES


Eric Jones, Travis Oliphant, Pearu Peterson, et al. {SciPy}: Open source scientific tools for {Python}. 


Jaume Garcia-Barnes, Debora Gil, and Aura Hernandez. Endowing canonical geometries to cardiac structures. In Oscar Camara, Mihaela Pop, Kawal Rhode, Maxime


REFERENCES


H. Hufnagel, X. Pennec, J. Ehrhardt, N. Ayache, and H. Handels. Comparison of statistical shape models built on correspondence probabilities and one-to-one correspon-
REFERENCES


REFERENCES


REFERENCES


D. Pinto, M.C. Robertson, P. Hansen, and J.H. Abbott. Economic evaluation within a factorial-design randomised controlled trial of exercise, manual therapy, or both


SEER. *SEER Training Modules - Structure of Bone Tissue*. National Cancer Institute, 12 June 2011.


REFERENCES


J. Wolff. The law of bone remodeling (original publication 1892 translated in 1986 by p. maquet and r. furlong), 1892.


