

A Novel Approach for the Visualisation and Progression Tracking of Metastatic Bone Disease

Nathan Timothy Sjoquist

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Nathan Timothy Sjoquist
December 2021

A Novel Approach for the Visualisation and Progression Tracking of Metastatic Bone Disease

Nathan Timothy Sjoquist

Abstract

Metastatic bone disease (MBD) is a common secondary feature of cancer that can cause significant complications, including severe pain and death. Current methods of diagnosis require a highly trained radiologist capable of interpreting medical images and recognising the sites of MBD. These medical images are often noisy, two dimensional, greyscale and usually have a poor resolution.

In order to help assist with these issues, several studies have shown that computer aided methods can locate MBD within medical images. However these methods are limited in scope, accuracy, sensitivity, explainability and do not improve upon the poor visualisations of the underlying medical imaging data.

To address these limitations, I have developed a novel method of automatic MBD assessment and visualisation using computed tomography (CT) imaging data as the input. The method is fully automated and does not require any human interaction – although users can interact with a viewer that visualises the results. This method has been tested on CT data from prostate cancer patients as prostate cancer is one of the most common sources of MBD.

The method described in this thesis has a sensitivity of 0.871 when detecting sclerotic and lytic lesions within a single data set. This sensitivity is comparable to existing methods, however the scope in detecting these lesions was limited to the vertebrae in previous studies. My method significantly expands this scope to include the ribs, vertebrae, pelvis and proximal femurs.

The work in this thesis also provides novel visualisations of the disease and does not suffer from explainability issues that plague modern machine learning algorithms.

In addition, I developed a novel method of tracking the spread of MBD at multiple time points using longitudinal CT data. This method is capable of calculating the change in lesion volume size across multiple time points, providing a novel numerical assessment.

I would like to dedicate this thesis to Amy. . .

Acknowledgements

I would like to acknowledge and to thank the Armstrong Trust for funding my studies.

I would also like to thank my family for their love and support – Steve and Darla Sjoquist, Matt and Nancy Bernard, Christina and Stephen Medawar, Mark and McKayla Bernard, Timothy Sjoquist, Laura Fairbrook, Allen Sugiyama and Kim Ulrich. I would particularly like to thank my father (Steve) for encouraging my interests in engineering and in computer science. And also my daughter Emma for being pure joy.

I would also like to thank my friends both in Cambridge and at home who have greatly enriched my life. A special thanks to Phil and Nancy Payne, Kennedy and Anthony Sanderson, Kate and Tyler Denmead, Joanna Burleigh, David Bedford, Robin Craster, Hugo Hadfield, Ryan Greenhalgh, Andrej Vasilj, Slawomir Tadeja, Cam Roberts, Juan Sebastián Cañavera-Herrera, Christian Buckingham and Dante McGrath. I am particularly grateful to Dante for proofreading this thesis. Furthermore, I would especially like to thank Brendan Payne for his friendship during this time.

Thanks also to my collaborators at Addenbrooke's Hospital – Ken Poole, David Thurtle and Vincent Gnanapragasam for helping me secure the CT data used in this project. Additionally, I would like to thank my advisor Andrew Gee, for sharing his knowledge about image registration, for providing constructive critical feedback on my work, and for examining this thesis. Further thanks to Marc Modat for also examining this thesis.

I am deeply grateful to Tristan Barrett, my secondary supervisor, for providing his medical expertise and for marking the lesions in the data set used in this thesis. And a very special thanks to Graham Treece, my primary supervisor, for his expert knowledge, guidance, encouragement and for his unwavering support for me.

Finally, I would like to thank my wife Amy Sjoquist, who encouraged me to pursue a PhD, who helped me through the difficult times, who travelled far from home to live in Cambridge, who has proofread and understood all of my research and who has been by my side every step of the way. I would not have finished without your encouragement. Thank you.

Table of contents

List of figures	xv
List of tables	xxv
Nomenclature	xxix
1 Background	1
1.1 Metastatic bone disease	1
1.2 Existing methods of diagnosis	4
1.3 Computed tomography	7
1.4 Existing detection methods	8
1.5 Existing methods of metastatic bone disease tracking	12
1.6 Existing visualisation methods	13
1.7 Cortical bone mapping	13
1.8 Research overview	16
2 MBD Assessment	19
2.1 Introduction	19
2.2 Segmentation of bone in CT	21
2.2.1 Overview of segmentation techniques	21
2.2.2 Hysteresis thresholding	29
2.2.3 Polygonal mesh background	29
2.2.4 Polygonal mesh generation	31
2.2.5 Bone thickness calculation	33
2.3 Atlas creation	34
2.4 Registration	38
2.4.1 Overview of mesh to mesh registration	38
2.4.2 Rigid registration	39
2.4.3 Deformable registration	40

2.4.4	Initial placement in registration	42
2.4.5	Piece-wise hierarchical registration	43
2.4.6	Improving the hierarchical registration	53
2.5	Symmetrical disease assessment	58
2.5.1	Smoothing of CBM data	59
2.5.2	Density differences	61
2.5.3	Diseased side selection	63
2.5.4	Visualisations of the disease	64
2.6	Longitudinal disease assessment	65
3	Results	69
3.1	Introduction	69
3.2	CT Data	69
3.2.1	Acquisition	69
3.2.2	Reference data	70
3.2.3	Disease shadow size	75
3.2.4	Converting disease shadow points to surface area	77
3.3	Evaluation of individual sub-methods	79
3.3.1	Hysteresis segmentation accuracy	79
3.3.2	CBM considerations	81
3.3.3	Validation of registration	81
3.4	Experiments on symmetrical disease assessment	86
3.4.1	Defining metrics	86
3.4.2	Overall results	86
3.4.3	Parameters	87
3.4.4	Determining parameters	88
3.4.5	Parameter considerations	91
3.4.6	Overfitting	92
3.4.7	Parameters based on volume size	93
3.4.8	Overall optimised results	94
3.4.9	Sclerotic and lytic labelling	106
3.4.10	Diseased side selection	114
3.4.11	Experiments using line data	116
3.4.12	Experiments on atlas symmetrical deformation mappings	117
3.5	Longitudinal evaluation	117
3.6	Running times	123

4	Discussion	125
4.1	Introduction	125
4.2	Experimental data	125
4.2.1	Segmentation results	127
4.2.2	Registration results	131
4.2.3	Parameter choice	134
4.2.4	Discussion of overall results	136
4.2.5	Discussion on volume calculation	137
4.2.6	Disease side selection	138
4.3	Discussion on general functionality	139
4.3.1	Usefulness pertaining to metastatic bone disease	139
4.3.2	Longitudinal analysis – changes in trabecular bone density over time	140
4.4	Visualisations	141
4.4.1	Use in a clinical setting	142
4.4.2	Clinical pilot study	143
4.4.3	Limitations	143
5	Conclusions	147
5.1	Introduction	147
5.2	Discussion and conclusions	147
5.3	Improvements	148
5.3.1	Improvements to segmentation	148
5.3.2	Improvements to registration	149
5.3.3	Improvements to CBM	149
5.4	Future Extensions	150
	References	153

List of figures

1.1	A computed tomography scan of the pelvis showing a sclerotic lesion in a prostate cancer patient. The lesion on the right side has a much higher bone density (shown by a more intense white colour in the image) than the healthy left side does. Image from the Addenbrooke's Hospital CT Data Collection.	2
1.2	Bone composition of the proximal femur showing the relative anatomy between the cortical bone and trabecular bone. The bone marrow (not shown for simplification purposes) is encapsulated within the trabecular bone. . . .	3
1.3	Imaging of MBD using a bone scan and plain X-rays. Images were adapted from [118].	4
1.4	A MRI scan showing a MBD lesion located within the right proximal humerus. Image adapted from [118].	5
1.5	A PSMA-PET scan showing uptake consistent with a bone metastasis. This image is from the Addenbrooke's Hospital Data Collection.	6
1.6	An example CT scan of the chest and spine. This image is from the Addenbrooke's Hospital CT Data Collection.	6
1.7	An example CT scan of the pelvis and spine. This image from the Addenbrooke's Hospital CT Data Collection.	7
1.8	An example of a CT sinogram. This image is from the Biomedical Imaging Group, EPFL [143]. Each pixel in this image shows the attenuation along a particular line: the vertical axis is the position and the horizontal axis is the angle of the line.	8
1.9	An example of a resliced CT image created by interpolating through a stack of CT images. This image is from the Addenbrooke's Hospital CT Data Collection.	9
1.10	General approach to MBD classification.	9

1.11	An example of measuring the cortical thickness along the cyan line through the cortical and trabecular bone. This image is a screenshot taken from Stradview [133]. The underlying CT image is from the Addenbrooke's Hospital CT Data Collection.	14
1.12	A graph of measuring the bone density by modelling the cortical thickness. Image is adapted from [135].	15
2.1	An example of plotting the HU overlap of non-bone and bone. Data was taken from a single full body CT data set. Although bone tissue has a higher radiodensity than non-bone tissue, there can be a significant overlap between the two in a single voxel due to the partial volume effect.	20
2.2	A CT scan of the pelvis. Bone segmentation in CT images can be difficult as bony regions and boundaries are not always well defined. The white arrows point to areas of bone that are difficult to segment. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.	21
2.3	An example of a typical segmentation with the segmented bone being outlined in purple. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.	30
2.4	This figure shows what a typical fully automatic hysteresis segmentation of the bone looks like, as well as indicating the typical range over which the detection of MBD is attempted. This is the result of the segmentation before the atlas registration. This image is a screenshot taken after the new mesh was loaded back into Stradview [133].	32
2.5	The pelvic bone is extremely thin in places, as can be seen in this CT image. The bone at these locations does not contain any trabecular bone but is entirely cortical bone. Underlying CT data is from the Addenbrooke's Hospital CT Data Collection.	33
2.6	A screenshot taken from Stradview [133] showing one of the atlases. Every point in the polygonal mesh has a perfectly symmetrical point pairing. The colour indicates the individual labelling of bones and bony regions.	35
2.7	A demonstration of joining the seam between severed triangles when creating the perfectly symmetric atlas. The two triangles $\triangle ORS$ and $\triangle SRT$ on the left are cut (indicated roughly by the red lines) by the plane of symmetry. New triangles $\triangle OO'S$ and $\triangle SO'S'$ are created to join the seam and complete the fully symmetric atlas.	37
2.8	An example of joining the centre seam. This process was used to help create a fully symmetrical atlas.	37

2.9	An overview of the registration method.	43
2.10	This image is an example of typical initial placement where the atlas is multicoloured and the target segmentation is coloured yellow. Initial placement in registration is very important as a poor initial placement can lead to inaccurate registration.	44
2.11	This image shows another example of a typical initial placement.	45
2.12	A diagram of the hierarchical registration. Initial registration first occurs globally with the entire skeleton. Progressively smaller regions are then registered with each previous step acting as the initial placement for the current step. Finally, each bone is deformed to match the segmentation. . .	46
2.13	An example showing a typical piece-wise hierarchical affine registration process of registering the atlas mesh to the target mesh.	47
2.14	An example of comparing a target segmentation to a deformed atlas segmentation. The target segmentation often contains a number of holes as seen in Figure 2.14a. This figure shows a typical example of the target segmentation mesh result of part of the pelvis. In Figure 2.14b, the deformed atlas was used to cover the holes.	48
2.15	The results of the piece-wise registration before the deformable registration has started.	49
2.16	The results of the piece-wise registration after 20 iterations of deformable registration.	50
2.17	The results of the piece-wise registration after 40 iterations of deformable registration.	51
2.18	In this image, the best affine registration is found, so the atlas (shown in purple), can best match the target (shown in yellow). The points in the atlas indicated by the white arrow are significantly far away from areas of the target mesh in comparison to the other points in the atlas. One such area is indicated by the black arrow. These points could easily pull the atlas in the wrong direction and so are ignored when computing the registration. . . .	52
2.19	A typical final registration – each region of the deformed atlas is coloured in a different colour while the target segmentation is in yellow.	53
2.20	An example of registering the lowest vertebra with the top of the pelvis. Using pre-marked points greatly improves vertebra registration.	55
2.21	An example showing the registration of an atlas vertebra to a target vertebra. Atlas vertebra A0 moves upward to register with target vertebra T0.	56

2.22	An example of using the top of a previously registered vertebra as the initial starting registration location for the vertebra above it. Atlas vertebra A1 is initially placed by aligning the $A1_{bottom}$ with the centroid of $A0_{top}$, $A0'_{top}$ and $A0''_{top}$	57
2.23	An example of a segmentation result overlaid with a colour map representing the trabecular bone density measurements. A user can easily see the difference in trabecular bone density by comparing the colour of the left side of the pelvis with the right side of the pelvis. This difference can indicate the presence of MBD.	59
2.24	An example of the custom image viewer. The right side of the image shows the main viewer with an alpha blend applied. This blends the polygonal mesh with the background and CT slice (texture mapped) image. The left side shows the CT slice 2D along with some of the user interface controls. This allows a user to specify the display options including the alpha blend, image brightness, image contrast, image zoom and slice image selection. . .	60
2.25	An example of the segmentation where each vertex has been coloured according to its trabecular bone density measurement.	61
2.26	An example of the segmentation where each vertex has been coloured according to its trabecular bone density measurement. This example used five iterations of smoothing.	62
2.27	An example of visualising disease by comparing trabecular bone density differences.	63
2.28	An example of re-slicing the CT data at a user selected location. A user can click on the mesh to calculate a re-sliced CT image at that location. Figure 2.29 shows this resliced CT image without the mesh. The CT data is from the Addenbrooke's Hospital CT Data Collection.	65
2.29	The diseased left side can be seen in the re-sliced CT image. The left side of the pelvis has a much greater trabecular bone density than the right side does. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.	66
2.30	In this example, the trabecular bone density from a baseline scan and follow-up CT scan has been projected onto the surface of their respective segmented meshes. By comparing these two images, it can be seen that the disease has progressed, as the trabecular bone density strongly differs in specific regions across these two images. The longitudinal disease assessment method attempts to find these differences.	66

2.31	An example showing the disease progression between a baseline case and a follow-up case. The disease progression (in turquoise) has been overlaid over the grey segmentation.	67
3.1	This slice image shows an example of a diseased region outlined by Dr. Barrett in red. Every lesion within each data set was outlined by expert radiologist Dr. Barrett. The lesion is in the left superior pubic ramus bone of the pelvis. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.	71
3.2	Another lesion outlined in red by the radiologist: this one is more challenging to find. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.	71
3.3	Visualisation of diseased 3D volumes created from the radiologist's 2D contour markings. Images are screenshots taken of the visualisation tool which was created for this project.	72
3.4	Location of lesions by region (rib, femur, etc.).	75
3.5	The diseased shadows on the segmentation mesh. The diseased vertices were found by intersecting each vertex normal on the segmentation with the triangles in the 3D lesions less than 10.8 mm away. The disease vertices are highlighted in fluorescent green (the red regions are protruding diseased regions). These projected diseased regions are called disease shadows. . . .	76
3.6	A plot showing the relationship between lesion volume and disease shadow size. As the volume of a lesion increases, so does the disease shadow size projected by that lesion. The fit (shown in red) can be modelled using a power of $2/3$ relationship. This is due to the conversion from volume in cm^3 to number points (which is a measure of area in cm^2). This fit was found using a least squares approximation method. A single 3D lesion can produce many disease shadows so the number of points on the y axis is actually the sum of the number of points in all of the disease shadows produced by that lesion.	78
3.7	An example of a segmentation (shown in yellow) that includes the CT table. The segmentation also extends far below the atlas femurs (shown in red and purple). Both of these issues can severely reduce the registration accuracy between the atlas (shown in multi-colour) and the segmentation. Note that in this case one of the ribs has also failed to register due to incomplete data in the segmentation.	83

3.8	The images from a case of an extremely diseased patient. The segmentation algorithm achieved poor results on this case as the spine, pelvis, and femurs contain very different HU values from that of a healthy CT scan. As this case would be very easy for a radiologist to find the lesions, it is not a good candidate for this algorithm and is not used in the experiments. The CT data is from the Addenbrooke's Hospital CT Data Collection.	84
3.9	These images show an example of how varying the threshold the HU difference cutoff can produce different results. The blue and red points are potentially diseased areas. The blue regions indicate positive differences greater than the cutoff value while red regions indicate negative differences greater than the cutoff value. Increasing the threshold HU cutoff will increase the amount of difference required in order for a vertex to be labelled as diseased.	87
3.10	These images show an example of how using a cluster size of 300 can be used to remove speckled noise.	88
3.11	These images demonstrate how using three smoothing iterations can remove speckled noise.	89
3.12	These images show an example of how using a number of smoothing iterations and a minimum cluster size can reduce speckled noise.	92
3.13	These graphs shows the relationship between sensitivity, volume and cluster size. In this experiment, the threshold cutoff size was set to 150, the number of smoothing iterations was set to 2 and the cluster size was allowed to vary. The lesions were binned into 20 groups by ranges of volume. As the cluster size increases, the sensitivity decreases.	95
3.14	A graph showing the cluster size vs point accuracy using the training data that does not include data from fold one. As the cluster size is increased, so does the overall point accuracy.	96
3.15	These plots show the relationship between sensitivity, volume and threshold. In this experiment, the cluster size was set to 7 and the smoothing to 2 iterations and the HU threshold cutoff was allowed to vary. The lesions have been binned by range of volume. As the threshold increases, the sensitivity decreases.	97
3.16	A graph showing the threshold size vs point accuracy using training data that does not include fold one data. As the threshold size increases, so does the overall point accuracy.	98

- 3.17 A plot showing the relationship between sensitivity, volume cutoff and cluster size. Lesions have been binned by their volume cutoff. Each volume is placed into every bin where its volume is greater than the threshold cutoff. For this experiment, the HU threshold was set to 150 and the number of smoothing iterations was set to 2. This graph is intended to show how each cluster size affects the sensitivity of detecting lesions that are greater than a certain volume. It does not show how a cluster size affects the sensitivity of detecting a lesion of a certain volume. Training data that does not include fold data one was used to produce this graph. 99
- 3.18 A plot showing the relationship between sensitivity, volume cutoff and threshold. Lesions have been binned by their volume cutoff. Each volume is placed into every bin where its volume is greater than the threshold cutoff. The cluster size for this experiment was set to 7 and the number of smoothing iterations was set to 2. This graph is intended to show how each threshold affects the sensitivity of detecting lesions that are greater than a certain volume. Training data excluding fold one was used to produce this graph. 100
- 3.19 A representative example of an average result from the fold five validation set. The assessment method found 5 out of 6 lesions correctly. It achieved a sensitivity of 0.833 with a point accuracy of 0.887. 101
- 3.20 A representative example of a good result from the fold five validation set. The assessment method found 23 out of 24 lesions correctly. It achieved a sensitivity of 0.958 with a point accuracy of 0.861. 102
- 3.21 A representative example of a good result from the fold five validation set. The assessment method found 23 out of 24 lesions correctly. It achieved a sensitivity of 0.958 with a point accuracy of 0.817. 103
- 3.22 A representative example of a good result from the fold five validation set. The assessment method found 19 out of 21 lesions correctly. It achieved a sensitivity of 0.905 with a point accuracy of 0.929. 104
- 3.23 A representative example of a poor result from the fold five validation set. This case achieved a sensitivity of 0.621 with a point accuracy of 0.823. The assessment method found 120 out of 193 lesions correctly. 105
- 3.24 The number of lesions grouped by region. The red bars show the number of missed lesions while the blue bars show the number of correctly detected lesions. This data is from the fold five validation data set. 111

3.25	An example of correctly labelling a lesion lytic. My algorithm labelled this lesion by symmetrically comparing the CBM trabecular bone density measurements. The CT data is from the Addenbrooke's Hospital CT Data Collection.	114
3.26	A example of incorrectly labelling a lesion as lytic. My algorithm labelled this lesion as lytic by symmetrically comparing the CBM trabecular bone density measurements. However, this lesion is actually sclerotic (although this has not been verified by a radiologist). This lesion was most likely labelled incorrectly as its symmetrically mapped points are also diseased and sclerotic as seen by the denser white bone on the left side of the image. The CT data is from the Addenbrooke's Hospital CT Data Collection.	115
3.27	A typical example of the disease progression. The bony segmentations have been displayed in grey while the lesions have been displayed in red. The disease has progressed from the baseline case to the follow-up case.	118
3.28	An example of the results from a longitudinal experiment. In the longitudinal assessment, strong differences in trabecular bone density between baseline and follow-up cases are marked as disease. These diseased regions are shown in teal. The baseline case had a total sum ground truth lesion volume of 120.845 cm^3 . The follow-up case had a total sum ground truth lesion volume of 264.592 cm^3 . This was an increase of 143.747 cm^3 . The predicted volume for this case was 101.442 cm^3 using the longitudinal assessment method. This is an absolute percent error difference of 29.431%	119
3.29	An example of longitudinal disease progression. It can be seen that the 3D lesions have significantly increased in number and in size from the baseline case to the follow-up case.	120
3.30	An example of a longitudinal result. The change in disease between baseline and follow-up cases is shown in teal on the bony surface. The baseline case had a total sum ground truth lesion volume of 1.249 cm^3 . The follow-up case had a total sum ground truth lesion volume of 305.602 cm^3 . This was an increase in 304.352 cm^3 . The predicted volume for this case was 150.501 cm^3 using the longitudinal assessment method. This is an absolute percent error difference of 50.6%.	122

3.31	This image shows the flow of the execution pipeline where each box is an executable. The output of each executable is used as the input of the next. The CBM and mesh bone generation executable and the MBD mesh generation executable can be run in parallel. Results can be calculated using the experimental results executable as well as viewed interactively in the 3D viewer or both.	124
4.1	Examples of typical diseased and healthy data sets. The healthy data sets extend to the shoulders, while the diseased data sets extend to the upper femurs.	127
4.2	A typical segmentation – the segmentation is not accurate in places. This is due to the partial volume effect, as both bone and non-bone is binned into the same voxel.	129
4.3	An example of a difficult region to accurately segment. Segmentation of the vertebrae is challenging due to the partial volume effect and the slice angle through each vertebra. An example of a typical vertebra in a CT scan – it is unclear where the upper segmentation boundary should be in this case due to the small angle difference between the surface of the vertebra and the CT slice. The CT data is from the Addenbrooke's Hospital CT Data Collection.	129
4.4	A typical poor quality segmentation result – the vertebrae are particularly problematic due to the extremely thin cortices.	130
4.5	A typical deformable registration result. The segmentation mesh is displayed in yellow and the individual pieces of the atlas are displayed by specific colours. The registration results are especially accurate within the pelvis and the femurs where the majority of the disease is located. It is important to note that there are some issues with the vertebrae registration.	132
4.6	The registration is not accurate in places as shown in this image. Significant differences in shape and size of the atlas and the segmentation can cause areas of poorly matching registrations.	133
4.7	An example of a poor symmetrical point match. This will cause a decrease in the overall accuracy of the algorithm, since clearly the wrong points are being compared. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.	133
4.8	A large 3D lesion (shown in red) produces a disease shadow (shown in green) with hundreds of points. This type of lesion is relatively easy to accurately locate as the cluster size can be set to a large number which removes much of the noisy error.	135

-
- 4.9 The very small 3D lesion (shown in red) produces a disease shadow (shown in green) with a tiny number of points. A very small disease shadow such as this is very difficult to distinguish from error. 136
- 4.10 An example of mapping the 3D visualisations back to the original CT images. On the left, a user has clicked on a diseased region (shown by the red surface normal line). On the right, this diseased data is shown in the resliced CT data (which is the plane defined by the vector found by averaging the red and blue surface normal vectors and by the vector of the green line connecting the two points). The CT data is from the Addenbrooke's Hospital CT Data Collection. Image is a screenshot taken of the custom image viewer. 142
- 4.11 The entire vertebrae shown in this image is diseased. The hardened trabecular bone density on each side cancels the other out so the region is not marked as diseased. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection. 144
- 4.12 The human skeleton does not develop in exactly a symmetric way. The bone within this CT slice is not perfectly symmetric, and this lack of symmetry is likely due to development as well as to disease. Additionally, the slice angle could be another cause of asymmetry with CT images. The CT data is from the Addenbrooke's Hospital CT Data Collection. 145

List of tables

2.1	Methods for full body bone segmentation in CT.	28
2.2	The numerical ordering (for clarity) of the vertebrae used in the registration.	54
3.1	The number of lesions binned into volume ranges. The majority of the lesions in this data set are less than 3.0 cm^3	73
3.2	The number of lesions binned into volume ranges less than 3.0 cm^3	74
3.3	The DSC metric and specificity for six segmentations. Specificity is the number of true positives divided by the (number of true positives plus the number of false positives). The DSC measurement is the ratio of correctly labelled voxels to total number of voxels.	80
3.4	The average surface distance metric for the six segmentations. The number of invalid points are the number of segmentation points that lie more than two standard deviations away from a ground truth point.	80
3.5	The average distance between each point in the segmentation and its closest respective point in the atlas, for 88 cases. The pelvic bone is made of the ilium, acetabulum, ischium and pubis bones.	85
3.6	The 10 best results and parameter values when sorting by point accuracy using a brute force method.	89
3.7	The 10 best results and parameter values when sorting by sensitivity using a brute force method.	90
3.8	The 10 best results and parameter values when sorting by point accuracy and sensitivity using a brute force method. The point accuracy and sensitivity were first normalised and then combined into a single term. These results were then summed and sorted.	91
3.9	The average results per case from using the optimal parameters when detecting data sets containing both lytic and sclerotic lesions using the validation data sets. 2 smoothing iterations, a cluster size of 7 and a threshold cutoff of 100 were used.	98

3.10	The average results per case from using the optimal parameters when detecting data sets containing both lytic and sclerotic lesions using the validation data sets. 2 smoothing iterations, a cluster size of 7 and a threshold cutoff of 125 were used.	98
3.11	The average results per case from using the optimal parameters when detecting data sets containing both lytic and sclerotic lesions using the validation data sets. 2 smoothing iterations, a cluster size of 7 and a threshold cutoff of 150 were used.	99
3.12	The sensitivity of detecting all lesions by volume (cm^3) using validation data from fold five.	107
3.13	The sensitivity of detecting lesions where the lesion volume was less than 3 cm^3 using validation data from fold five.	108
3.14	The sensitivity of detecting lesions by disease shadow size using validation data from fold five.	109
3.15	The sensitivity of detecting lesions by disease shadow size where the number of points is less than 300 using validation data from fold five.	110
3.16	The sensitivity of detecting lesions by disease shadow size for lytic lesions using validation data from fold five.	112
3.17	The sensitivity of detecting lesions by disease shadow size for sclerotic lesions using validation data from fold five.	113
3.18	The table shows the number of detected and missed lesions out of 1626 using line data sampled at 40 locations using 2 smoothing iterations and a cluster size of 7. The full data set was used in the experiment (instead of using a cross validation approach). Results would most likely be worse using CV approach.	116
3.19	The point accuracy using the optimised parameters described in 3.4.8 (using the parameters optimised by cross validation on the single data sets).	120
3.20	The 10 best longitudinal accuracies letting the HU threshold and the cluster size vary. Ranges were set near to the optimised parameters described in 3.4.7 (using cross validation to optimise the parameters on single data sets), although cross validation was not used to optimise these results (across multiple data sets).	121

3.21	The 10 best results by volume percentage difference. The volume percentage difference is the absolute value of the calculated volume subtracted from the ground truth volume divided by the ground truth volume. Initial ranges were set near to the optimised parameters described in 3.4.7 (using cross validation to optimise the parameters on single data sets), although cross validation was not used to optimise these results (across multiple data sets).	121
3.22	10 representative longitudinal cases showing total lesion volume (cm^3) of the baseline and the follow-up scans. Volume growth is the change in volume (cm^3) from baseline to follow-up scans. The volume estimate (cm^3) is the predicted volume growth which uses equation (3.3) to convert shadow points to volume. Percentage difference shows the relationship between the volume growth and the volume estimated growth.	123
3.23	The table shows running times for the different executables in the symmetrical assessment method.	123
3.24	The table shows running times for the different executables for the longitudinal assessment.	124

Nomenclature

Acronyms / Abbreviations

ASD Average surface distance

ASSD Average symmetric surface distance

CBM Cortical bone mapping

CT Computed tomography

CV Cross-validation

DICOM Digital imaging and communications in medicine

DSC Dice similarity coefficient

HU Hounsfield unit

MBD Metastatic bone disease

MRI Magnetic resonance imaging

PET Positron emission tomography

RECIST Response evaluation criteria in solid tumours

SPECT Single photo emission computed tomography

SVM Support vector machine

Voxel Volume pixel (a pixel with volume)

Chapter 1

Background

1.1 Metastatic bone disease

Metastatic bone disease (MBD) is cancer that has spread from a primary tumour into the bone [63] and is a common secondary feature of prostate cancer, breast cancer and many other malignancies [30, 35]. Bone metastasis is of great clinical importance in patients with these diseases, as MBD can cause considerable complications [30] that include pain, impaired mobility, pathological fracture, cranial nerve palsies, nerve lesions, hypercalcaemia and spinal cord suppression [19, 63, 100]. It is estimated that metastasis is the cause of 90% of cancer related deaths [113]. MBD is particularly common in patients dying of advanced cancer of the breast, bronchus, prostate, kidneys and thyroid [107] with the skeleton being the most common site of metastasis [35]. Patients with these types of cancers account for more than 80% of the incidence of MBD [35].

As these cancers are not common in younger patients, bone metastases are usually encountered in the middle-aged and the elderly, but MBD may occur in children as well [100]. Of these cancers, prostate cancer is one of the most prevalent, and is the second leading cause of malignancy in men [18, 103].

Each year, in the United Kingdom, almost 50,000 new cases of prostate cancer are diagnosed [142] and 15,000 men die from its complications [107]. Studies show that prostate cancer patients have a MBD incidence approaching 70% [19, 30, 35, 63, 107] in advanced cases with an estimated incidence of 80-100% in men where the cancer is terminal [63]. Because prostate cancer is so common and because it produces a high incidence rate of MBD, my thesis will primarily use data from prostate cancer patients to further research in this field, although this research could be applied to other types of primary cancer.

The majority of skeletal metastases are haematogenous in origin as the primary cancerous tumour cells travel through the arteries and veins [100]. In normal bone, development and

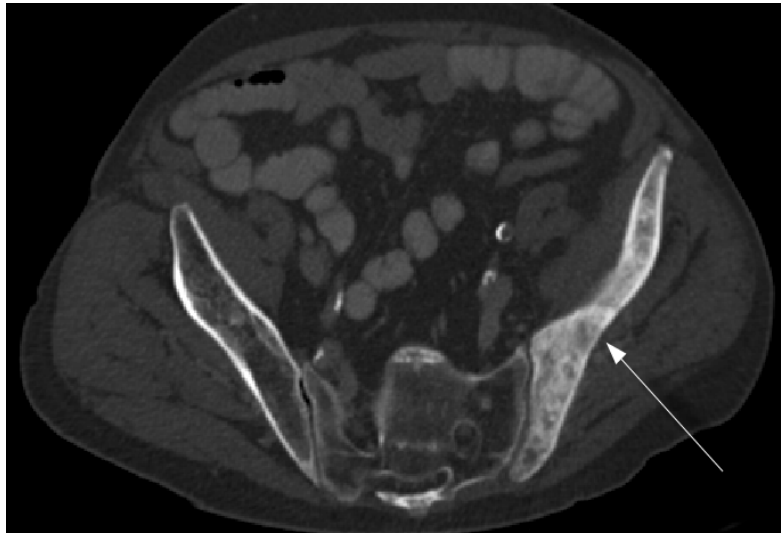


Fig. 1.1 A computed tomography scan of the pelvis showing a sclerotic lesion in a prostate cancer patient. The lesion on the right side has a much higher bone density (shown by a more intense white colour in the image) than the healthy left side does. Image from the Addenbrooke's Hospital CT Data Collection.

maintenance of bone tissues are sustained through a balance of osteoclasts and osteoblasts resorbing and depositing bone tissue. Osteoclasts, attracted to a site of fatigue damage, remove the fatigued bone by creating an erosion cavity [30]. Osteoblasts are then attracted to sites of prior resorption, and synthesise an organic matrix that will fill the resorption cavity [30]. With MBD, this process of bone resorption and formation is disrupted by the cancerous cells, which cause metastatic bone lesions to form [35]. These metastatic bone lesions can be of three types: sclerotic, lytic and mixed [35, 63, 100]. Most bone metastases due to prostate cancer are sclerotic, causing an increase in bone density as shown in Figure 1.1, although some metastases are lytic, causing a reduction of bone density [63] as well as mixed (lytic/sclerotic). Mixed can also refer to the data set itself if it contains both sclerotic and lytic lesions. In all cases, MBD causes an abnormal change in the density of bone.

Two types of bone exist as seen in Figure 1.2, a harder outer shell called the cortical bone and a softer inner spongy bone which encapsulates the bone marrow, called the trabecular bone [37]. The softer trabecular bone has a much greater surface to volume ratio than the harder outer cortical bone, and consequently, metastatic involvement of the trabecular bone surfaces are often more developed [30, 35].

MBD most often occurs in the five years following treatment of the original cancerous tumour [100]. MBD usually forms at multiple lesion sites [35, 63] in the axial bones, such as the vertebral column, ribs, skull as well as in the pelvis and femurs [30, 35, 63, 107, 114]

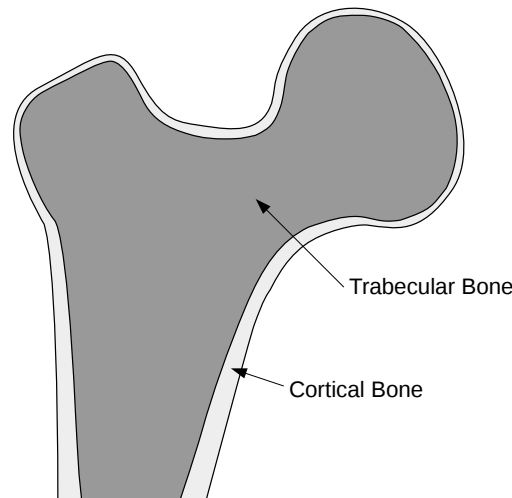


Fig. 1.2 Bone composition of the proximal femur showing the relative anatomy between the cortical bone and trabecular bone. The bone marrow (not shown for simplification purposes) is encapsulated within the trabecular bone.

while bone metastases of the distal bones are rare although not unheard of [30, 35]. MBD lesions almost always cause pain and this pain is usually the presenting symptom for a patient [30, 35]. This pain is caused by a variety of factors including periosteal stretching, compression or infiltration of nerve roots, reflex muscle spasm and the local effects of cytokines [35]. Bone pain is also the most common source of pain in patients with cancer, with over two-thirds of cancer patients having pain severe enough to require treatment at some point, much of this being caused by MBD [19, 63].

In general, metastatic bone disease is an incurable condition [35, 100]. However, it is still important to diagnose MBD early in order to provide treatment and improve quality of life. Treatment of bone metastasis is primarily palliative [100] in order to relieve pain, prevent development of pathological fractures, and to improve mobility and function [100]. It is also sometimes possible to prolong survival by preventing the metastatic tissue from spreading as well as to reduce organ destruction [35, 100]. Common treatments include surgery, radiotherapy, endocrine therapy, chemotherapy as well as pain relieving drugs [19, 30, 63, 100, 114]. It is also imperative to diagnose MBD because it may lead to a diagnosis of the primary cancer that was previously undiagnosed [35]. It may also indicate the staging of the cancer [35], which has a profound effect on both the treatment and prognosis. There is however, probably a considerable underestimate of the true incidence of metastases in the skeleton [107].

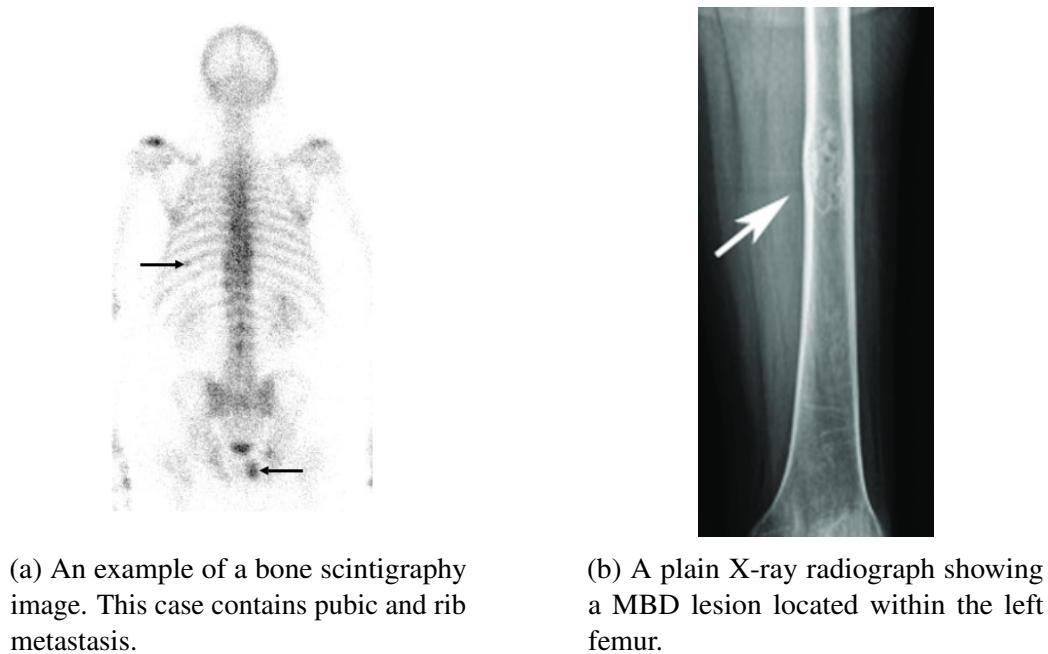


Fig. 1.3 Imaging of MBD using a bone scan and plain X-rays. Images were adapted from [118].

1.2 Existing methods of diagnosis

Imaging is used as a pivotal tool in diagnosing MBD [100]. Imaging allows a radiologist to recognise the development of new MBD lesions as well as to track the enlargement of existing lesions – both of which indicate the progression of the disease [35]. The diagnosis of MBD is usually straightforward, but MBD can occasionally be confused with benign skeletal pathology [63]. In some cases of MBD, the disease is very difficult to identify [100] as it can be challenging to interpret radiographic changes [35].

The usual imaging techniques for MBD include bone scintigraphy (Figure 1.3a), radiographic films (Figure 1.3b), computed tomography (CT) (Figure 1.1) and magnetic resonance imaging (MRI) (Figure 1.4) [30, 35, 63, 100, 107].

More modern MBD imaging techniques include single photon emission computed tomography (SPECT) [102] and positron emission tomography (PET), which is shown in Figure 1.5 [31]. Bone scintigraphy remains the most widely used method in the diagnosis and surveillance of metastatic skeletal disease [63] and is the usual method of choice for physicians [107]. CT can also be effectively used as a problem solving modality for diagnosing MBD [114] and can be combined with SPECT and PET to attain more functional imaging information, due to uptake of the radioactive tracers [31, 102]. Evaluation of patients using MRI is also useful, especially as the response to therapy can be easily followed

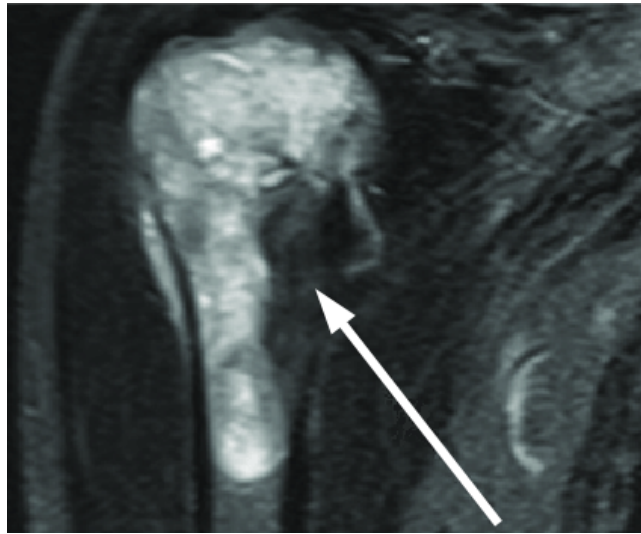


Fig. 1.4 A MRI scan showing a MBD lesion located within the right proximal humerus. Image adapted from [118].

using MRI [114]. However, the CT imaging modality is much more commonly used than MRI [116].

Whilst CT is not the most common way to diagnose MBD, in this thesis I will be focusing on this modality for a number of reasons. Although planar scintigraphy is the most widely used imaging tool in diagnosing MBD, it is more useful in screening MBD than in finding exact bone lesion sites as it is only in 2D [52]. Plain radiographs can show MBD, but are often difficult to interpret and again are only in 2D [30]. A significant amount of bone destruction must exist before either bone scintigraphy or radiographs show the abnormality [114], while CT can detect small areas of bone destruction [151]. A MRI scan is twice as expensive as CT [52] and is far less available in hospitals according to a study in the United States [47]. PET is a very expensive imaging modality [52] and the effectiveness of both PET and SPECT in diagnosing MBD is still not proven, as very few studies use these modalities for MBD [52].

CT provides excellent high definition bone imaging and is especially useful in imaging metastatic bone lesions [89]. CT is the most widely used imaging technology in radiology departments [23] and is available in most hospitals [47]. It is also cheaper than most other imaging modalities [52]. Radiation dose is also less of a concern in patients with advanced cancers. CT is also used routinely for oncology follow-up in the majority of cancer cases, and provides a good assessment of bone, lung and visceral metastases.

Furthermore, the use of CT enables the use of the relatively new cortical bone mapping (CBM) technique (described in Section 1.7). Whilst CBM has been tried on MRI as well as on CT, it was specifically designed for determining bone quality in CT data and has only been

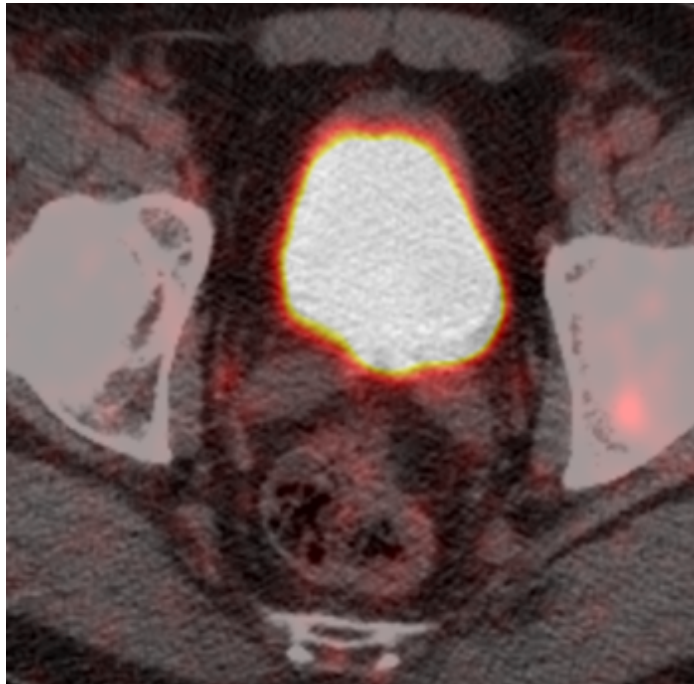


Fig. 1.5 A PSMA-PET scan showing uptake consistent with a bone metastasis. This image is from the Addenbrooke's Hospital Data Collection.



Fig. 1.6 An example CT scan of the chest and spine. This image is from the Addenbrooke's Hospital CT Data Collection.

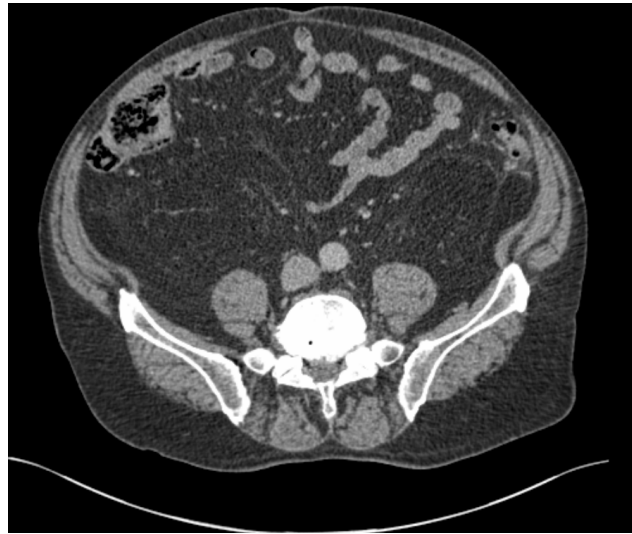


Fig. 1.7 An example CT scan of the pelvis and spine. This image from the Addenbrooke's Hospital CT Data Collection.

thoroughly validated on this modality. This technique allows bone density to be accurately estimated in 3D, unlike plane radiographs. CBM has never been used to detect bone lesions, however it is anticipated that this detailed analysis of bone quality could be an important factor in assessing MBD.

1.3 Computed tomography

X-ray computed tomography is an imaging modality that produces cross-sectional images as seen in Figure 1.6 and in Figure 1.7, representing the X-ray attenuation properties of the body [122]. X-rays are produced by an X-ray tube, attenuated by the patient, and measured by an X-ray detector [122]. As the X-rays travel through the patient, different types of materials attenuate (by absorbing and scattering) the X-rays by different amounts based on the material's properties. This produces varying X-ray measurements depending on the patient's internal material makeup. Narrow beams of X-rays, placed in a fan or parallel beam geometry, scan over a field of view densely sampling the patient [23, 122]. This is repeated over many angles and often at many z-depths to create a full scan of an area. A sinogram is a 2D scan of an area at a given z-depth [23], an example of which can be seen in Figure 1.8.

Based on these measurements, the actual attenuation at each point of the scanned slice is reconstructed using a direct Fourier reconstruction or a filtered back projection technique [23]. Reconstruction can also be performed by using an iterative process that uses both forward (from the reconstruction to the sinogram) and backward (from the sinogram to the recon-

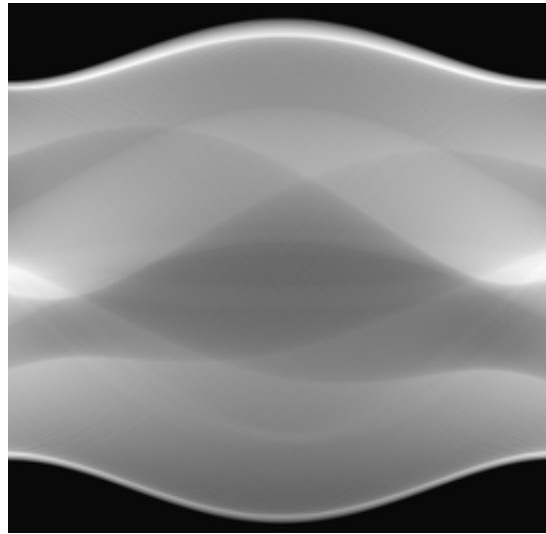


Fig. 1.8 An example of a CT sinogram. This image is from the Biomedical Imaging Group, EPFL [143]. Each pixel in this image shows the attenuation along a particular line: the vertical axis is the position and the horizontal axis is the angle of the line.

struction) steps to refine the reconstruction. The attenuation values are usually represented as greyscale voxel values and are normalised by their relationship to the attenuation of the value of water and also of air [122]. This is called a Hounsfield unit (HU) with -1000 HU being the value assigned to air and 0 HU being the value assigned to water. The scale is open ended, but many digital images only store values up to 3000 HU [23]. Stacks of images can be resliced at any location and angle by interpolating through the images (as seen in Figure 1.9) and can be of great value to radiologists [114]. While 3D visualisations (usually volume rendering) are very impressive, radiologists tend to rely on either the original images or on resliced images when diagnosing.

1.4 Existing detection methods

There are a small number of existing methods that use computer-aided detection to semi-automatically or automatically find the presence of MBD. Most of these methods use the CT imaging modality [22, 25, 53, 55, 60, 61, 99, 105, 106, 115, 141, 149, 154–156, 162, 163] although a few methods use MRI [64, 147]. One method uses PET and CT in combination [161] and one method uses bone scintigraphy [28]. Some of these methods detect only lytic lesions [64, 115, 149, 162, 163] or sclerotic lesions [22, 105, 106, 141, 155, 156] although some methods can find both [53, 55, 60, 61, 99, 161]. These methods are limited to



Fig. 1.9 An example of a resliced CT image created by interpolating through a stack of CT images. This image is from the Addenbrooke's Hospital CT Data Collection.

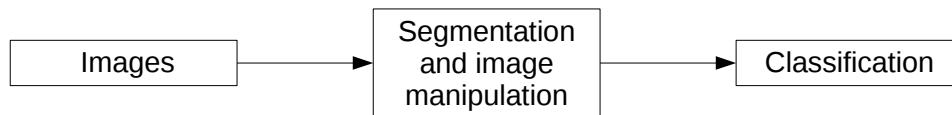


Fig. 1.10 General approach to MBD classification.

spinal data except for only one study that uses the full body, but in this case the analysis is restricted to 2D data [28].

All these methods generally follow a similar approach, which is shown in Figure 1.10. The bone is first located in the images using a semi-automatic or fully-automatic segmentation method. By segmenting the bone, the search space is narrowed (since by definition, MBD will only appear in or near the bone), which reduces the complexity of the problem. The segmentation methods used include thresholding [22, 60, 106, 155, 161–163, 156], model and atlas based [25, 28, 55, 64, 99, 115], marginal space learning [53] and manual selection [147, 149]. Often, these methods are combined with other image manipulation techniques such as region growing, graph cuts or the watershed algorithm, which are used to merge or prune

regions in order to improve segmentation accuracy and to manipulate the morphology of the segmentation [22, 61, 106, 155, 156, 162, 163].

After the bone is located, various approaches are used to locate and label the diseased bone within each bony region. The simplest method involves using the cross-section of the segmented bone to create a histogram by binning voxels based on HU [55]. The histogram is compared to a healthy region's histogram where strong differences in histogram shape indicate lytic or sclerotic lesions [154]. However, this method uses a manual segmentation approach that is tedious and time consuming. Manual segmentation also limits this method's usefulness as it can only be used to distinguish between healthy and unhealthy regions but cannot be used to find the location of the regions themselves.

More complex supervised learning methods have also been employed to locate and label the diseased bone. In supervised learning, an unknown function is estimated given a labelled training set [91]. Each data point in the training set is made of a number of features which describe that data point. Once the function is estimated by learning from the training set, the estimated function can make predictions to generalise the output of the original unknown function given novel inputs [91]. The methods for detecting MBD that use these supervised learning approaches can be described as classifiers as they attempt to associate a class to each novel data point.

Classifiers are dependent on training sets comprised of meaningful features as they must learn to accurately classify healthy bone and diseased bone based on the training data. Many of these approaches use hand-picked MBD image features such as shape, density and location [156, 162] as well as mean intensity [99], Haar-like features [53, 149], and others [61]. Medical experts are needed to mark the data (i.e., healthy or diseased) in order to provide the classification for the training data. This marking can be very time consuming. Types of learning classifiers used in this way include support vector machines [22, 155, 156, 161–163], Fisher's linear discriminant [64], graph cut classification [61], Markov random field in the Bayesian classification framework [99] and random forests [53, 149]. Neural network classifiers that do not require hand-picked features have also been used [25, 60, 105, 106, 147].

It is difficult to compare the accuracy of these methods with each other as the input data varies significantly from study to study. These differences include using completely different imaging modalities as well as having different types of markings produced by the experts. Within the data itself, small indistinct MBD lesions are much more difficult to locate than large distinct lesions. This further complicates the comparison of methods as information describing lesion data is not often available, making it challenging to judge how difficult the lesions are to actually find. Also, the final classifications (diseased, abnormal, healthy,

sclerotic, lytic, etc.), groupings (voxel based, object based, region of interested based, etc.) and reporting (sensitivity, accuracy, false positives, etc.) vary significantly between studies.

In the study that produced the best results (by sensitivity and number of false positives), a convolutional neural network was used in conjunction with a medial axis transform and a random forest to classify spinal lesions [25]. The study used CT data consisting of thirty-one cases, involving thirteen females and eighteen males which contained a total of 626 vertebrae [25]. Lesions (both lytic and sclerotic) were marked by expert radiologists, voxel-wise, and divided into three categories, depending on volume, ranging from small (1.40 mm^3 to 14.20 mm^3) to medium (14.2 mm^3 to 300.00 mm^3) to large (greater than 300.0 mm^3) [25]. Connected voxels were grouped into objects for testing purposes [25]. The classifier performed a voxel-wise classification and results were determined by comparing the overlap of the classified voxels to marked voxel object groups [25]. Using an object overlap intersection threshold of 1%, the method produced a 0.92 to 0.95 true positive rate for small lesions, 0.97 to 0.99 for medium lesions and 0.99 for large lesions [25]. This method produced a mean sensitivity of 0.88 when detecting lytic lesions and a sensitivity of 0.97 when detecting sclerotic lesions. It also produced 45.63 to 130.31 false positives per patient for small regions, 20.47 to 40.40 false positives per patient for medium regions, and 5.92 to 7.65 false positives per patient for large regions [25].

The study that produced the second best results (by sensitivity and number of false positives) used a support vector machine to identify MBD lesions in a PET/CT spine data set consisting of 26 diseased cases and 18 healthy cases [161]. Radiologists identified 456 lytic and sclerotic lesions which were manually labelled and used for training and testing purposes [161]. Using a cross-validation approach, the method achieved a sensitivity of 0.81 in identifying lytic regions with 1.3 false positives per case, a sensitivity of 0.81 in identifying sclerotic regions with 2.1 false positives per case and a sensitivity of 0.76 in identifying both lytic and sclerotic regions with 2.1 false positives per case [161]. However, this method also produced an exponential increase in a number of false positives for small improvements in the sensitivity [161]. Information about the size of the diseased regions was not included in this study, which is important as larger volumes are almost always easier to detect than smaller regions. The definition of sensitivity is also not discussed, so it is unclear how much of a detected region needs to overlap a marked region in order to classify the marked region as detected. It appears that a single overlapping point is enough to mark the lesion as detected. This method also requires PET data along with CT data to achieve its sensitivity, which is not desirable.

The study with the next best results (by sensitivity) used a convolutional neural network and a random view aggregation to detect spinal sclerotic metastases in CT images consisting

of forty nine patients (14 female and 35 male) [105]. Radiologists labelled a total of 532 lesions, all of which were greater than 300 mm³ [105]. At three false positives per patient, the method produced a sensitivity of 0.70 in detecting sclerotic regions [105]. However, this method also produced an exponential growth of the number of false positives with slight increases to sensitivity [105].

These methods are limited in scope as they can only find MBD in the vertebral column. This is a severe limitation as MBD is often located within the pelvis, ribs and upper femurs as well as in the vertebral column [30, 35, 63, 107, 114]. The single method that works on the whole body [28], is underdeveloped as it is only a segmentation method and is limited to 2D planar bone scintigraphy data. Furthermore, these methods produced a high number of false positives that severely reduces their usefulness. Also, many of these methods either do not perform well on small lesions or do not describe how well they perform on small lesions. This is a vital metric in the usefulness of these methods, as detecting MBD early (when the lesions are small) helps improve the effectiveness of the treatment.

There is cause to suspect that, however well these supervised learning techniques work on recorded vertebral data, this performance would not carry through to general whole-body data. This is because there is significant variance in bony shape and size when comparing the spine to other parts of the full-body. Supervised learning also often requires large quantities of accurately marked training data, which can be difficult to obtain in a medical imaging setting due to patient ethical agreements and a lack of experts to mark the data [26]. It is also currently very difficult to understand and explain why a supervised learning method classifies a data point to a particular class [57, 58]. This makes the justification of the classification often impossible, which can be detrimental in a medical setting [58]. Supervised training methods can also suffer from over-fitting where noise and variations in the training data are modelled instead of the true signal [91]. This can also be caused when enough of the true signal is missing from the training data set, if the training data does not contain an accurate representation of the real data. Over-fitting can cause the classifier to fail when it encounters data that does not fit the observed training patterns. Again, as medical data for training is difficult to attain, it is likely that the training set will not contain enough data to learn all of the features it needs to accurately locate MBD lesions in all patients scanned in a hospital setting.

1.5 Existing methods of metastatic bone disease tracking

It is also of great importance to track metastatic disease progression by imaging the patient at different time-points in order to monitor the disease spread so the patient can receive optimal

treatment. Without automated tools, this is a slow tedious process that involves comparing longitudinal images side by side to look for small bony changes.

A limited number of studies have attempted to address this problem using a temporal subtraction technique in which a baseline image is automatically compared to a follow-up image through image subtraction [94, 108, 141]. These methods attempted to map the bony voxels between the CT data sets by registering one with the other. This was accomplished by warping or deforming one data set (in either 2D or 3D) with the other to account for differences in body position between the two scans.

While these methods are an important proof of concept, they go no further than demonstrating their potential usefulness in helping improve the performance of radiologists. These methods could also be improved to provide a much better visualisation of the disease as they currently only provide 2D slice difference images. 2D images still will take a long time for a radiologist to examine and to interpret. Also, these methods do not numerically summarize and quantify the changes in diseased volume between scans. This is a currently unavailable but very useful and desired metric.

1.6 Existing visualisation methods

Methods for detecting and tracking MBD do not generally describe how their results are visualized but instead provide a few examples where the lesions were marked on the original 2D images. Visualisation of disease is important, as much of medical imaging only produces 2D greyscale images that can be difficult to interpret. Much of the research in this thesis will focus on producing better and unique methods of visualising MBD.

Very few MBD visualisation methods exist other than viewing the imaging scans directly. However, a method was developed that can automatically segment the bones in a thoracoabdominal CT and then erode the cortical bone, providing an unobstructed view of the trabecular bone [132]. However, this method does not highlight the diseased regions, so expert knowledge from a radiologist is still required to interpret the images.

1.7 Cortical bone mapping

A more accurate way to locate and measure cortical bone and trabecular bone is the cortical bone mapping technique. Given that MBD affects bony features, detecting MBD in CT data critically involves the ability to accurately assess bone quality from such data. CBM is an automatic method of estimating bone quality, focusing on the cortical and sub-cortical trabecular bone, and primarily aimed at the CT imaging modality [136]. It can produce

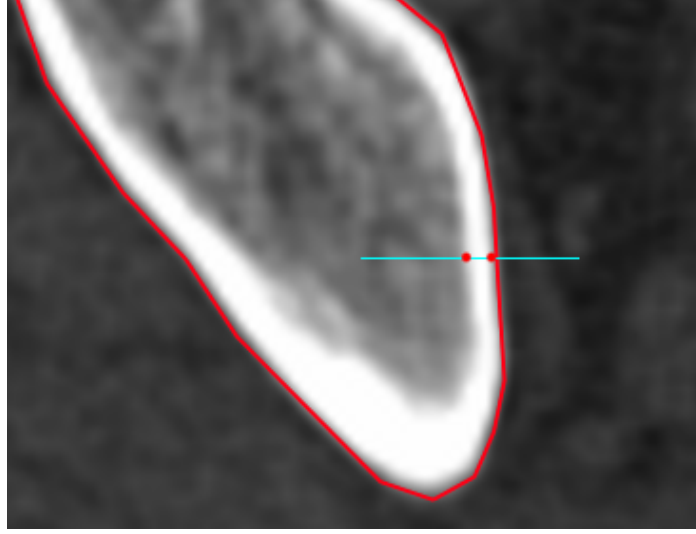


Fig. 1.11 An example of measuring the cortical thickness along the cyan line through the cortical and trabecular bone. This image is a screenshot taken from Stradview [133]. The underlying CT image is from the Addenbrooke's Hospital CT Data Collection.

cortical thickness measurements down to 0.3 mm from clinical CT as well as estimate cortical and trabecular bone density accurately. It has been used successfully in a number of studies to assess bone integrity [134–137]. CBM fits a parameterized model to measurements taken along a short line (usually approximately 18 mm) perpendicular to the cortical layer as seen in Figure 1.11 [136].

The underlying density distribution is assumed to be equation (1.1), where y_0 , y_1 and y_2 are the CT values of the surrounding tissue, cortical and trabecular bone respectively, t is the cortical thickness and $H(x)$ is a unit step function that models the location of the cortical and trabecular bone which can be seen in Figure 1.12 [137].

$$y(x, t) = y_0 + (y_1 - y_0)H(x + \frac{t}{2}) + (y_2 - y_1)H(x - \frac{t}{2}) \quad (1.1)$$

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}} \quad (1.2)$$

$$h(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sigma\sqrt{2}} \right) \right] \quad (1.3)$$

To account for the imaging system's blur, the system's point spread function can be approximated as a Gaussian function shown in equation (1.2), where σ represents the extent of the blur [101, 135–137]. This is a deliberate simplification of the real point spread function and, partly as a result, the value of σ is allowed to vary locally over the image data. The

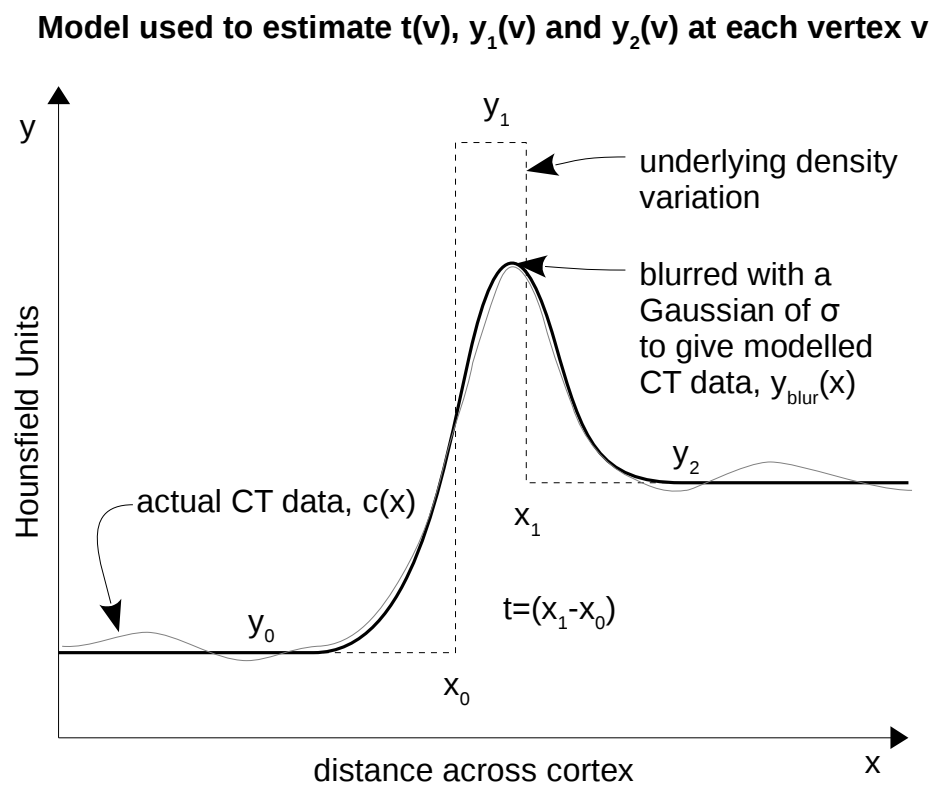


Fig. 1.12 A graph of measuring the bone density by modelling the cortical thickness. Image is adapted from [135].

step response of the imaging system is equation (1.3) where erf denotes the error function that is the result of integrating the Gaussian in equation (1.2). Equations (1.1) and (1.3) are combined together to produce the final blurred density function shown in equation (1.4).

$$y_{\text{blur}}(x, t) = y_0 + \frac{(y_1 - y_0)}{2} \left[1 + \text{erf} \left(\frac{x + \frac{t}{2}}{\sigma\sqrt{2}} \right) \right] + \frac{(y_2 - y_1)}{2} \left[1 + \text{erf} \left(\frac{x - \frac{t}{2}}{\sigma\sqrt{2}} \right) \right] \quad (1.4)$$

An optimisation method is used to optimise equation (1.4) to fit this function to the sampled CT data: the Levenberg-Marquardt [90] approach has been shown to work well on this problem [135]. The optimisation is ill-posed in that there are multiple similar solutions. In particular, for thin cortices, the thickness t and the cortical density y_1 can vary with respect to each other without much impact on the result. To force convergence, y_1 is estimated and constrained which reduces the unconstrained number of parameters down to five [135–137]. In order to map the cortical and trabecular bone thicknesses across an entire surface, the outer cortical bone surface is located and encapsulated within a triangle mesh [134]. Measurements are taken at the vertices within this mesh, with their corresponding surface normals providing the orientation of the cortical bone [136].

1.8 Research overview

In this thesis, I describe a new method that can find these bone lesions by symmetrically comparing (left to right) the bone density in the human skeleton and grouping areas of large difference. I also describe a new method that can track lesion growth using multiple time-points providing an automatic numerical comparison as well as improved visualisations. These methods are thoroughly tested to demonstrate their validity.

The goal of this thesis is to improve upon existing methods for the detection and tracking of MBD. Where existing methods are limited in scope to the spine, this work expands detection to the ribs, spine, pelvis and femurs as MBD is commonly found in all of these areas. The current supervised learning methods provide unexplainable results with high false positive rates and poor visualisation of disease. This work provides much better visualisations that can explain the results and help reduce or explain the false positives. It will use CT imaging as this modality is readily available in most hospitals and provides high definition bone imaging able to display small and developing MBD lesions essential to diagnosing MBD at an early stage. The work in this thesis will also improve upon MBD tracking as current methods are underdeveloped and suffer from poor visualisations.

The current clinical methods of diagnosing MBD lesions from CT data require time consuming reviews of many cross-sectional images looking for small and diffuse bony changes – the summary of which is difficult to demonstrate in a multidisciplinary team meeting review setting. These changes are also often difficult to demonstrate to the patient. The aim of this work is to address these challenges, by developing a novel method that can more thoroughly identify these lesions to better detect and visualise MBD.

Rather than using training data to learn features distinct to bone lesions, the CT data is compared to itself through symmetry to find irregularities. This removes the need for training data and bypasses the problem often encountered with machine learning algorithms – poor results when the training data does not contain a specific, never-before-encountered pattern.

Using symmetry to locate disease has been shown to be a successful approach in a variety of situations [129, 160]. Although results from these papers could be improved, they demonstrate the feasibility of the approach. I have extended these ideas and have applied them to the problem of locating MBD.

The human skeleton is nearly symmetric in both shape and bone density, and does not change over time as rapidly as MBD would be expected to. When two healthy symmetric bony regions are compared with each other, the regions will hence contain little difference. However, MBD causes sclerotic and lytic bone lesions to form, increasing the density (sclerotic lesions) or decreasing the density (lytic lesions) of the bone. When the bone density of symmetric regions containing bone lesions is compared, the difference is greatly increased as the lesions do not generally form symmetrically across both regions.

These lesions (areas of differing densities) are displayed through a novel visualisation. These visualisations have the potential to improve accuracy in diagnosis of MBD and to reduce the time-consuming review of the individual inspection of hundreds of cross-sectional images. This visualisation technique could be used to identify potential areas of concern, although the final reporting of a bone lesion would require assessment of the source data by the radiologist.

Additionally in this thesis, I explore a novel way of tracking MBD progression in patients imaged (using CT) at multiple time-points. Instead of using symmetry in a single scan, the trabecular bone density is compared between longitudinal scans to find differences or the lack thereof. This can be used to indicate disease progression and to assign a quantifiable value to the progression.

This new method is developed in Chapter 2, the experimental results are located in Chapter 3, these results are discussed in Chapter 4, and Chapter 5 contains concluding remarks.

Chapter 2

MBD Assessment

2.1 Introduction

As metastatic bone disease is located within the skeleton, finding all areas of bone within each CT image is an essential step in its detection and tracking. In CT bone segmentation, voxels are classified as being either of a bone type or of a non-bone type. Voxels labelled as bone generally contain a higher intensity than that of non-bone (unless the CT data contains a metal implant or a contrast agent). Cortical bone has a radiodensity typically greater than 1000 HU while soft tissue and fat have a radiodensity typically between -70 HU and 100 HU. However, in the region between 100 HU and 1000 HU, there is much overlap between the radiodensity of trabecular bone, fat and soft tissue as seen in Figure 2.1. This is mainly due to the partial volume effect where multiple types of tissue contribute to the average HU value recorded for a single voxel.

Because of this, a single intensity threshold cannot be used to accurately separate bone from non-bone, as a threshold set low enough to include all bone will also include fat and soft tissue. A threshold set high enough to exclude fat and soft tissue will also exclude bone. To overcome this issue, context from surrounding voxels must be combined with voxel intensity to correctly classify a voxel as being either bone or non-bone.

Furthermore, although the more dense cortical bone encapsulates the less dense trabecular bone, the boundary between the two is not always clearly defined in CT images, as seen in Figure 2.2. This is due to a combination of the partial volume effect and the imaging blur resulting in limited imaging resolution, as well as complications from MBD, as lesions can affect both the cortical bone and trabecular bone morphology.

Segmentation is further complicated by the great variance in shape, thickness and size between different types of bones as well as by the significant variance in anatomy and body position between patients. Since correct classification of bone voxels is not straightforward,

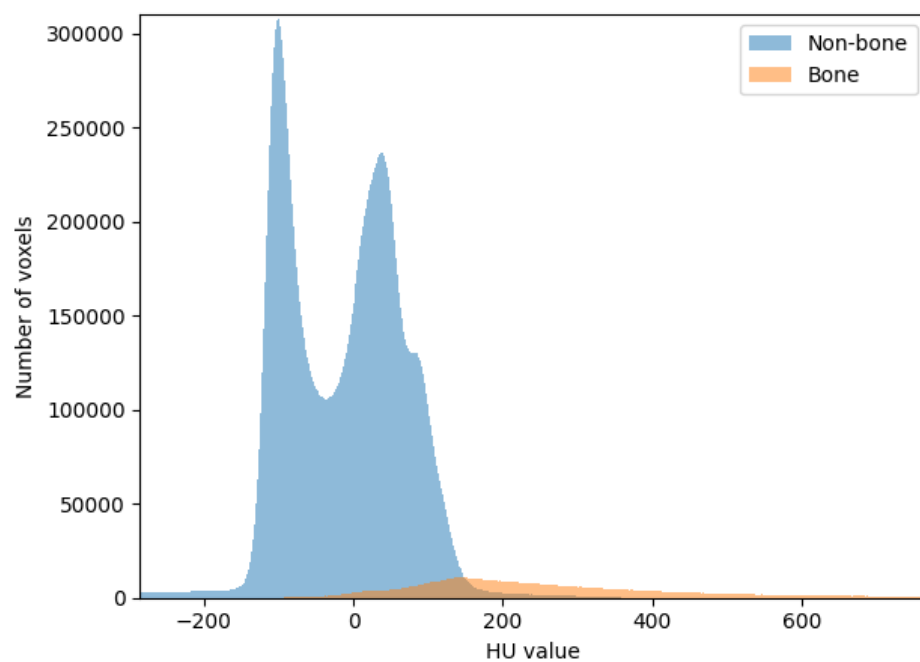


Fig. 2.1 An example of plotting the HU overlap of non-bone and bone. Data was taken from a single full body CT data set. Although bone tissue has a higher radiodensity than non-bone tissue, there can be a significant overlap between the two in a single voxel due to the partial volume effect.

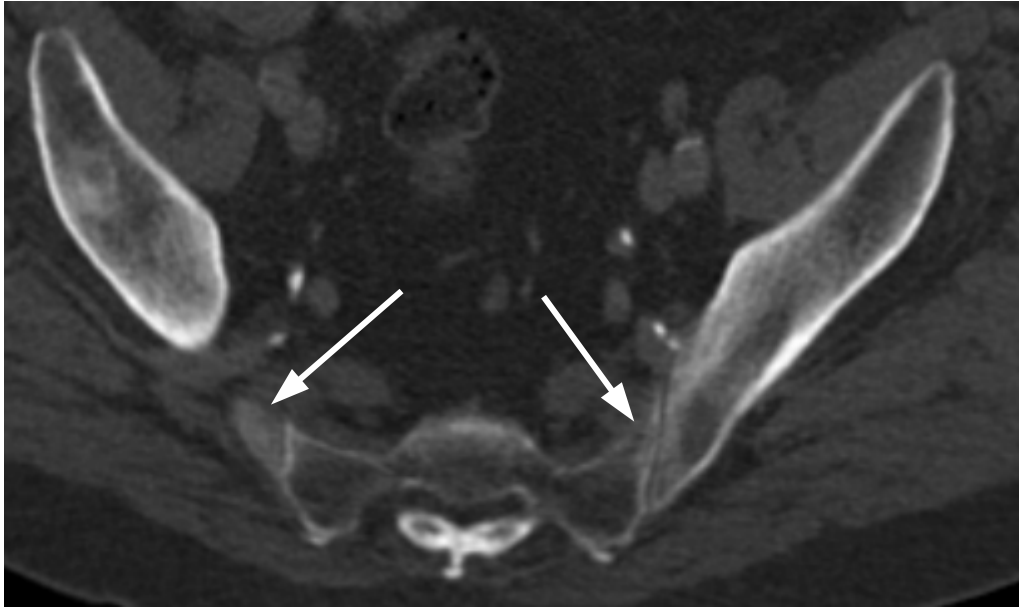


Fig. 2.2 A CT scan of the pelvis. Bone segmentation in CT images can be difficult as bony regions and boundaries are not always well defined. The white arrows point to areas of bone that are difficult to segment. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.

and the large variation in anatomy and pathology make developing good priors difficult, the segmentation of bone in CT is a challenging problem.

2.2 Segmentation of bone in CT

2.2.1 Overview of segmentation techniques

Many approaches have been developed that can segment bone in CT. The ideal method for detecting and tracking MBD needs to be capable of segmenting most of the body (at least the chest, abdomen and pelvis including ribs, spine, pelvis and upper femurs), have a high accuracy rate (especially the bone surface boundaries), require no human interaction (as manual bone segmentation is very time consuming even for a small section of the body) and be relatively quick to run on the resources that are available.

To determine segmentation accuracy, segmentation results are usually compared to a ground truth segmentation that has been carefully segmented through a manual process. Comparison between segmentation methods is often difficult as unique CT data sets are used in each study and individual studies often use different comparison metrics. These comparison metrics include the dice similarity coefficient, average surface distance, average

symmetric surface distance, volume overlap error, relative volume difference, mean absolute error, Hausdorff distance, average Hausdorff distance, Jaccard similarity and others, or just involve visual comparisons.

The most commonly used similarity metric in CT bone segmentation is the Dice Similarity Coefficient (DSC). The DSC metric (equation 2.1), initially developed by Sørensen et al. [119] and Dice [34], can be used to compare two discrete sets of data.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (2.1)$$

In equation (2.1), X and Y each contain a discrete set of elements and $|X|$ and $|Y|$ are the cardinalities of the two sets. When applied to segmentation methods, one set (i.e. X), contains the ground truth voxel classification of bone or non-bone. The other set (i.e. Y), contains the segmentation method's voxel classification results. A higher DSC score means more voxels have correctly overlapped between the segmentation and the ground truth data sets.

Another common approach used in determining the segmentation accuracy, is to compare the 3D surface representations of the segmentation and ground truth data sets by measuring the average surface error between the two point sets. There are two commonly used similarity surface metrics – the average surface distance (ASD) and the average symmetric surface distance (ASSD).

$$ASD = \frac{1}{|Y|} \sum_{y \in Y} D_X(y) \quad (2.2)$$

$$ASSD = \frac{1}{|X| + |Y|} \left(\sum_{x \in X} D_Y(x) + \sum_{y \in Y} D_X(y) \right) \quad (2.3)$$

The average surface distance is shown in equation (2.2), while the average symmetric surface distance is shown in equation (2.3). In these equations, Y is the set of points that comprise the surface segmentation and X is the set of points that comprise the ground truth surface. $D_X(y)$ is the distance from a point (y) in Y to its closest point in X or vice versa if Y and X are switched. The average surface distance is a measure of how close the segmentation surface is to the ground truth surface, while the average symmetric surface distance is a more robust measure of how close the segmentation surface is to the ground truth surface and also how close the ground truth surface is to the segmentation surface.

CT segmentation approaches can be organised into general techniques that include global thresholding [21, 46, 54, 123–125, 144, 145, 148], adaptive thresholding [17, 65, 150, 166, 168], hysteresis thresholding [43], deformable models [72, 96, 110, 111], region-

growing [65, 84], watershed [40, 165], active contour [97, 140], edge based [148], level set [85], graph cuts [44, 75, 93], statistical shape models [112, 128] and machine learning [7, 14, 27, 38, 45, 56, 62, 71, 76–79, 82, 109, 131, 157].

Most of these methods are only designed to segment a specific bone or a small group of bones, which is a far easier task than segmenting all the bone within the full body. Methods have been designed to segment the skull [65], orbital bone [77, 78], temporal bone [82], mandible [128, 131, 144, 145], radius [21, 44], ulna [44], carpal bones [110, 111], distal bones [45], ribs [17, 84, 157], vertebrae [6, 7, 38, 40, 62, 72, 79, 84, 85, 93, 97, 166], pelvis [56, 109, 112, 165], hip [27, 123, 150, 168], femur [65, 75, 124, 150], knee [65, 96, 125], tibia [21, 96, 148] and fibula [96].

In many of these segmentation methods, manual human interaction is required to produce acceptable results. Examples of these include Brahim et al. [17], Buie et al. [21], Patrick and Indu [97], Pardo et al. [96], Sebastian et al. [110, 111], Truc et al. [140] and Yu et al. [165]. In one common approach (as in Buie et al. [21]), a global threshold was manually set to segment the data. Individual images were then manipulated to improve results (using techniques such as dilation and erosion in order to fill surface gaps). Another common approach that required manual input (as in Sebastian et al. [110, 111], Pardo et al. [96] and Truc et al. [140]) used a region growing or active contour method in which a region or contour was expanded or shrunk to lie upon the bony region or bony boundaries. Because the initial placements for these methods are difficult to automate in an accurate way within relatively complex CT data, these approaches require manually placed initial starting seeds or starting contours for each 2D bone region. As a fully automatic approach is needed, the methods described in this paragraph are not appropriate for the requisite segmentation.

Selecting a segmentation method that produces high accuracy is important for this research. However, a number of the segmentation methods cited above produced very poor results and thus can be ignored for this purpose as not being sufficiently useful. This is especially true for techniques that use global thresholding in which a single threshold is selected to separate bone from non-bone based on voxel radiodensity.

Of all the studies that use global thresholding, van Aarle et al. [145] produced a method that achieved the best results (to the best of my knowledge). They developed a way to optimise the global threshold to minimise classification errors using information from both the raw DICOM data and the reconstructed images. Although this method has better accuracy than all other global thresholding methods, it still produced a significant number of misclassified voxels, so a more complex segmentation technique is required than global thresholding.

A number of methods have been developed that produce moderately accurate results. Gassman et al. [45] used an artificial neural network and achieved segmentation to ground

truth overlap of 76.0%. Li et al. [82] achieved a DSC of 77.18% by using a deep learning method. Krčah et al. [75] used graph cuts and boundary enhancements and reported that 81.0% of cases were accurately segmented. Zhang et al. [166] used 3D adaptive thresholding to achieve a volume overlap between the segmentation and ground truth of 85%. Nguyen et al. [93] used 3D graph cuts and achieved a volume extraction accuracy ranging from 82.30 to 86.37%. Although these methods segment a significant amount of bone correctly, a number of segmentation methods produced much better results with an accuracy of greater than 90.0% [14, 27, 43, 56, 65, 71, 125, 128, 150, 168].

While discussing these methods, it is important to note that even methods that produced high accuracy are often not good segmentation methods for detecting MBD for various other reasons. Westin et al. [150] used a tensor based adaptive filtering approach that produced a 98% voxel overlap between the segmentation and ground truth. Although this method produced very good results, it was only tested on a single case and on a specific bony region (upper femur and a small section of the pelvis). As the testing of this method was extremely limited, it is difficult to predict whether this method would perform well when segmenting the full skeleton in a much larger CT data set. Furnstahl et al. [44] used a graph cut approach to segment the radius and ulna in CT. The surface results from this approach were compared to a ground truth surface that was manually segmented. It produced an average surface distance of 0.22 mm. However, this method used a cost function that was tailored specifically to segment the radius and ulna and will not likely apply well to the segmentation of a whole CT skeleton.

Kang et al. [65] used 3D region growing with local adaptive thresholds, border closing and boundary adjustment for segmenting femurs, skulls and knees. This method produced segmentation specificity errors of less than 1% in finding bone volume. However, the study stated that this approach performed poorly on diseased subjects, so it would be unwise to use this method in detecting MBD. Seim et al. [112] used a statistical shape model to segment the pelvic bones in CT. The model deformed while respecting shape constraints to form the segmentation. This method produced an average surface distance of 0.7 ± 0.3 mm. However the study also stated that it does not handle pathological data well, which makes it a poor choice for segmenting data containing MBD.

A number of methods produced very accurate results but are tailored specifically to individual bones (as are all of the other methods mentioned above). Chu et al. [27] used a statistical shape model with a random forest for segmentation and produced a DSC of 93.9%, 96.5% and 96.4% for the pelvis, left femur and right femur respectively. Although this method produced good accuracy, it would be difficult to apply a single statistical shape model of the entire human skeleton to a CT set as there is much more variability in the shape

and size of a full skeleton when compared to that of a single bone. Hemke et al. [56] used a convolutional neural network to segment the pelvis and produced a DSC of 92.0%. Sun et al. [125] used an active contour with prior prediction information to segment the knee. It produced an average segmentation DSC greater than 94%.

Taghizadeh et al. [128] used a statistical shape model and template matching correction to segment the scapula, mandible and orbital bones. It was tested on 63 scapula, 72 mandible and 50 orbit data sets and produced an average DSC of over 90%. Zhou et al. [168] used an adaptive thresholding approach and then refined the border using surface normal information. It was tested on 70 hip joints and produced a Jaccard similarity measure of 96%.

There are also a number of methods that are tailored specifically for segmenting individual vertebrae in the spine. Liu et al. [85] used a level set method to segment the vertebrae, achieving a DSC of 94.8 to 97.7%. Lessmann et al. [79], Bae et al. [7] and Janssens et al. [62] used a convolutional neural network that produced segmentation DSC accuracies of 94.7%, 95.2% and 95.8% respectively. Dutta et al. [38] used a U-Net to segment vertebrae and achieved an accuracy of 96.0%.

While the methods mentioned above produced good results and have helped further the field of bone segmentation in CT, it is unclear what segmentation accuracy these methods would produce when applied to the full skeleton. It is also unclear how some of these methods *could* actually be applied to the full skeleton as they have been designed to segment specific bones. As there is no evidence for the performance or applicability of these methods on the whole body, they will not be pursued further in this thesis.

Four methods exist that *have* been designed to segment bone within a full body CT scan. Bieth et al. [14] used an iterative random forest approach using Harr-like features as well as geometric and distance features. A hierarchical label tree was used to find bone centroids in order to obtain the geometric and distance measurements within bony regions. The CT data was re-sampled to an isotropic resolution of 2 mm and an image size of 230×230 voxels. This method was tested on twenty full body CT scans of healthy subjects as well as on thirty prostate cancer and twenty myeloma data sets of the thorax and trunk. It achieved average DSC scores of 84.4%, 81.6% and 74.8% respectively [14]. It used a 2-fold cross validation approach for training and validation. The average segmentation time for one scan was about 12.48 minutes.

This method appears to be a viable full body segmentation technique, although it produced a somewhat low DSC and the paper does not include a discussion from where the main sources of error stem. This is important as certain bony regions are more difficult to segment than others making comparisons between other methods difficult. The re-sampling of the image size might also be problematic for MBD visualisations, as the standard CT image

size is 512×512 voxels. Resizing would lose resolution that might be important for the detection of small MBD regions.

Klein et al. [71] used a U-net convolutional neural network adapted from Ronneberger et al. [104]. This method was tested on 18 full body CT data sets of patients suffering from myeloma. It achieved a DSC of 95.0% using a 6-fold cross-validation approach. The method used images of axial, sagittal and coronal slices to try and take into account 3D information. After training, each scan took about 9 minutes and 30 seconds to segment. Best results were achieved in the femurs, followed by the pelvis and upper body while the ribs produced the poorest results. This method produced a specificity of 0.94 ± 0.04 and a sensitivity of 0.91 ± 0.08 .

The DSC from Klein et al. [71] is better than that of Bieth et al. [14] although this method was only tested on eighteen data sets. This method segmented standard sized 512×512 CT images which would make visualisation of disease easier. The results were only tested on data sets containing myeloma so it is not clear how this method would fare using a data set containing healthy bone or a data set containing MBD. Overall, this could be a valid method of segmenting full body CT.

Fu et al. [43] performed bone segmentation by optimising the morphology of a 3D anatomical atlas mesh to match a target segmentation that was segmented using hysteresis thresholding. Hysteresis thresholding addresses the problem of overlapping regions of bone and non-bone by provided context that greatly improves the accuracy when compared to global thresholding.

In hysteresis thresholding, multiple thresholds are used and combined to locate the bone voxels. These bony voxels comprise their own segmentation (which I will refer to as the target segmentation). An anatomical surface mesh atlas was then registered (point mesh to point mesh) to the outer surface voxels of the target segmentation. Registration was performed by following a hierarchical registration tree that involved registering the atlas to the target. Starting with the full atlas, at each step of the tree, smaller and smaller sections of the atlas were registered with the target. Each previous registration step helped guide the current registration into the correct place. Finally, the atlas was deformed to match the target segmentation.

The deformed atlas surface was superimposed on the original CT data set and used to label individual voxels as being either bone or non-bone. The original target mesh was discarded as the final deformable atlas became the final segmentation. 19 patients scanned in an oncology department (with no mention of a specific disease) were used as a test set. This method produced a segmentation DSC of 90.0% [43].

This DSC is higher than in Bieth et al. [14] but also lower than Klein et al. [71]. However, there is reason to suspect that results produced by this method, have better surface accuracy than do Bieth et al. [14] and Klein et al. [71] as this method is specifically a surface segmentation based method. This method also produced an average surface distance measurement between the final segmentation surface and the ground truth segmentation surfaces [43]. Results were tabulated by individual bone groups and range from 0.41 ± 0.02 mm to 1.55 ± 0.20 mm. The number of test cases (19) was comparable to the other methods of Bieth et al. [14] (30 full body cases) and Klein et al. [71] (18).

There is little discussion in this paper about bony areas that achieve poor results [43]. There is mention, however, that sometimes the vertebrae may be misaligned due to significantly different skeleton shapes and so must be guided by human intervention for correction [43]. This is undesirable behaviour as the MBD detection and tracking algorithm should be fully automatic. This method took around twenty minutes per scan to segment with the deformable registration taking almost half of this time [43] – longer, though not significantly, than the first two methods. It also used the standard 512×512 sized CT images, which would enable better visualisation of MBD.

In the final method, Lay et al. [76] used a watershed algorithm in conjunction with support vector machine (SVM) classifiers to segment bone in full body CT. First, the data was heavily smoothed, then the watershed algorithm was used to find general bone surface boundaries. These boundaries helped guide the decomposition of CT images into supervoxels [76]. These supervoxels were merged with each other depending on proximity and similarity. A SVM was used to classify voxels into being either bone or non-bone. A second SVM classifier was used to classify supervoxels that were difficult for the first classifier to classify in order to improve accuracy. Further SVM classifiers were sometimes also used for more specific regions of the body to improve results. This method was trained on 57 scans and then validated on 80 CT scans (collected from hospitals without mention of a specific disease). It achieved a DSC of 98.2% with a specificity of 0.986 and a sensitivity of 0.979 [76]. Each CT scan segmentation took around two to three minutes making this method the fastest of the four methods. There was no mention of a specific CT image size used in this paper [76]. There was also no discussion about specific bony regions that were difficult for this method to segment.

Of these four methods (summarised in Table. 2.1), Lay's method (watershed and support vector machine) achieved the best overall results with a DSC of 98.2% [76] compared to Bieth (random forest) with a DSC of 84.4% [14], Fu (hysteresis thresholding and hierarchical registration) with a DSC of 90.0% [43] and Klein (U-net type convolutional neural network) with a DSC of 95.0% [71]. Of these methods, Lay et al. [76], Bieth et al. [14] and Klein et al.

Table 2.1 Methods for full body bone segmentation in CT.

Author	Method	DSC
Lay et al. [76]	Watershed and support vector machine	98.2%
Klein et al. [71]	U-Net convolutional neural network	95.0%
Fu et al. [43]	Hysteresis thresholding and hierarchical registration	90.0%
Bieth et al. [14]	Random forest	84.4%

[71] used machine learning approaches, while Fu et al. [43] did not. All of these methods ran within an appropriate time allotment. However, it is significant that none of these papers mention if these methods have been tested on CT data sets containing MBD.

It is also significant that only Klein et al. [71] and Lay et al. [76] mentioned specific specificity and sensitivity rates. These are important metrics, as the DSC is a measure of total overall accuracy (bone and non-bone) while specificity and sensitivity is a measure of bone accuracy. As there is generally much more non-bone than bone in a CT image, it is possible for the segmentation method to produce a high DSC in segmenting the non-bone correctly while still producing poor results in segmenting the bone. Fu et al. [43] however, while not including specificity and sensitivity metrics, produced average surface distance metrics that are very helpful in determining the accuracy of the bone surface.

Although Lay's method produced a higher DSC when compared to the other methods, it is not clear if this would translate to better results in detecting and tracking MBD. In my research, I used CBM, which accurately measures trabecular bone density, useful in locating MBD. CBM produces the best results when the contour lines surrounding the bone lie exactly between the bone and non-bone pixels, though it will still produce results when the boundary is between 1 or 2 mm from this. It is essential for the bone boundary to be as accurate as possible as this boundary also determines the surface normals along the bone surface, critical to producing accurate CBM modelling. Due to the heavy smoothing, it could very well be that Lay's method achieves the best overall accuracy in finding the majority of the volumetric bone correctly, but produces poorer surface boundary results. When examining the images in Lay's paper [76], multiple regions of the bony surface are clearly incorrect, which is likely to be problematic. The only method that used a surface based segmentation optimisation method is Fu [43].

As none of these methods have been tested on data sets containing MBD, it is unclear whether a machine learning approach would produce good results when segmenting bone that contains patches of MBD lesions, as these patches could vary greatly from the normal bone training data. Klein's method [71] worked successfully in segmenting bone in CT of

myeloma patients, but it is unclear how much damage myeloma causes to bone in these data sets (MBD often causes a significant amount of bone damage).

2.2.2 Hysteresis thresholding

For this research, I implemented a hysteresis thresholding and hierarchical registration approach, which was based on Fu [43]. The hysteresis thresholding technique produces a relatively high DSC in segmenting bone in CT, but, more importantly, should produce better surface normals than the other full body methods as it explicitly optimises the surface boundaries to find the best segmentation. This method also produces an acceptable average surface distance metric demonstrating acceptable bone surface errors. In addition, hysteresis thresholding enables contour lines to be produced with sub-pixel accuracy by interpolating between thresholded voxels. This further improves subsequent CBM density calculation as the surface normals will be more accurate. However, segmentation is only a small part of the approach I outline in this thesis, and it is acknowledged that, whilst Fu's method seems to be the most promising, the other full body methods may also be appropriate for this task.

In the hysteresis thresholding implementation, all voxels with a radiodensity greater than 400 HU were labelled as bone. After this, voxels with a radiodensity greater than 200 HU and that were adjacent (eight-way connected) to the first set of voxels (those with a radiodensity greater than 400 HU) were also labelled as bone. Additionally, voxels were marked as bone if they had radiodensity greater than 200 HU and if any of the nine voxels (eight-way connected, as well as the centre voxel) in the slice directly below or above had a radiodensity greater than 400 HU.

A contour algorithm was then used to trace around each segmentation to find the border voxels. The contour line was adjusted to subpixel accuracy by weighting the difference between each bony voxel and its neighbouring non-bony voxel. The contours were saved and exported into Stradview for polygonal mesh surface generation and CBM.

2.2.3 Polygonal mesh background

There are a number of ways of rendering and modelling 3D data. General approaches include ray tracing [153], volume rendering [80], curve based (i.e. 3D spines) and polygonal mesh rendering [12]. (A variety of lesser used methods exist as well, such as radiosity based, photon mapping and others, but they are largely irrelevant).

The polygonal mesh was used in this project as it offers many benefits. The mesh is a way of defining the shape of a polyhedron through a collection of vertices (points in 3D space – usually with a colour and a normal vector), edges (connections between two vertices)

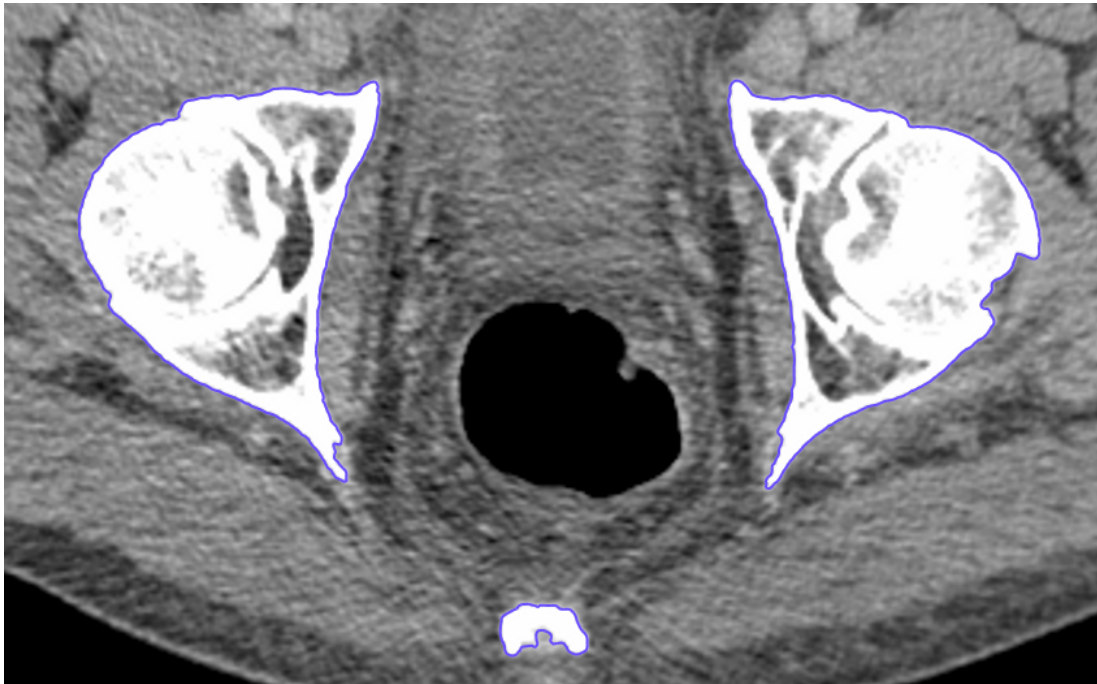


Fig. 2.3 An example of a typical segmentation with the segmented bone being outlined in purple. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.

and faces (a closed set of edges – usually triangles). This was first described by Baumgart [12] for use in computer vision algorithms and has been extended into mainstream use in the field of computer graphics. The polygonal mesh is the standard 3D surface format supported by many modern software packages (i.e. OpenGL [95]) as well as graphics hardware. 3D graphics cards have been designed and optimised to quickly and efficiently render large 3D polygonal meshes using fragment shaders. These cards also make use of efficient pixel shading programs that can quickly calculate semi-realistic lighting effects for each mesh vertex. This allows for state of the art visualisations and often real time rendering of interactive computer 3D graphics.

Polygonal meshes were also used in this thesis as they can be created at any resolution and can be used to precisely model the surfaces of 3D objects (i.e. bony surfaces). Meshes can be created in a variety of ways. Often polygonal meshes are created through interactive modelling software such as Maya [87], or by using a scanning based approach. Polygonal meshes can also be created from discrete scalar fields by using the marching cubes [86] or marching tetrahedra [36] algorithms.

The polygonal mesh is also an efficient way of storing 3D surface data. Efficiency was needed as the surface meshes used in this project are relatively large, consisting of hundreds of thousands of vertices. This information (along with the normal vectors, edge

relationships and triangle surface information) was needed to be able to fit into memory to support real-time viewer interactions.

Furthermore, the Stradview software [133] supports polygonal meshes. This software was needed as it contains the implementation of the CBM algorithm as well as the marching tetrahedra algorithm that was used to convert discrete segmentation contour points into a polygonal mesh.

Most of the other rendering methods mentioned at the start of this section make use of the polygonal mesh. In ray tracing, the scene is usually defined as a polygonal mesh. Rays are projected into the scene and intersections are found with the mesh faces (usually triangles). At the intersections points, the rays are split into reflection rays and refraction rays. The colour intensity is recursively summed in order to render the scene.

Although ray tracing can be used to create photo realistic renderings, it produces the most impressive results when the scene has either reflective or opaque objects within it. The bony meshes used in this thesis are neither reflective nor opaque, so using ray tracing would not have improved the visualisations. Also, ray tracing is computationally expensive and cannot often be rendered in real-time, a necessity for an interactive viewer.

Curve based methods also make use of polygonal meshes. Splines are usually interpolated into 3D points at a given resolution in order to create a polygonal mesh. The polygonal mesh is then rendered using a standard polygonal mesh rendering technique. Curve based methods were not used in thesis as the surface was already defined through the segmentation, so the surface did not need to be modelled using a curve based approach.

In volume rendering, volumetric data is usually displayed directly. However the data is sometimes converted into a polygonal mesh before rendering. Although volume rendering can produce interesting visualisations, it was not used in this project as this approach is very similar to displaying slice planes at various angles through the CT data, which was too limited for the purposes of this thesis.

2.2.4 Polygonal mesh generation

Once the bone in a CT volume was segmented using hysteresis thresholding, the 2D segmentation contours in each CT image slice were converted into a 3D triangle polygonal mesh which fully encapsulates the bone. Figure 2.3 shows what a typical segmentation bone outline looks like. The Stradview software implementation [133] was used to convert each cross section into the mesh using a shape-based interpolation technique [139] (to add more contour points) and the regularised marching tetrahedra technique [138] to generate the mesh. An example of a typical segmentation mesh using hysteresis thresholding can be seen in Figure 2.4, which shows the result before the atlas registration.

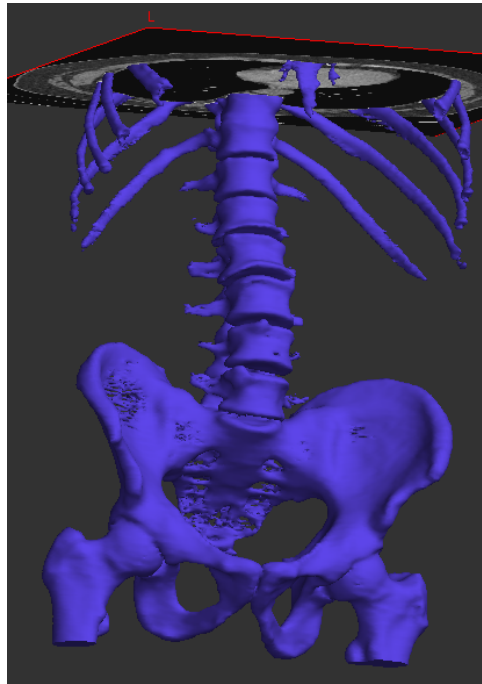


Fig. 2.4 This figure shows what a typical fully automatic hysteresis segmentation of the bone looks like, as well as indicating the typical range over which the detection of MBD is attempted. This is the result of the segmentation before the atlas registration. This image is a screenshot taken after the new mesh was loaded back into Stradview [133].

These operations were performed using function calls in the Stradview software package using a command line interface. These commands are documented on the Stradview website [133] on the scripts documentation page. The `SURFACE_RESOLUTION`, `SURFACE_SMOOTH_TYPE`, `SURFACE_SMOOTHING` and `SURFACE_ALL` functions were called to build the mesh with the resolution set to high, the surface smoothing type set to standard and the surface smoothing set to low. The `THICKNESS_EST_DENSITY`, `THICKNESS_SURF_FIT` and `THICKNESS_SURF_SAVE` functions were called to run CBM and to save the results.

The segmentation is correct and very accurate across most of the surface, although problems can clearly be seen by a visual inspection of the mesh. Specifically, this segmentation contains a number of holes, but the otherwise accurate surface locations mean that this segmentation is highly appropriate to detect, visualise and track MBD. The segmentation DSC results can be seen in section 3.3.1 and the overall results in detection of MBD using this type of segmentation were also recorded in section 3.3.1.

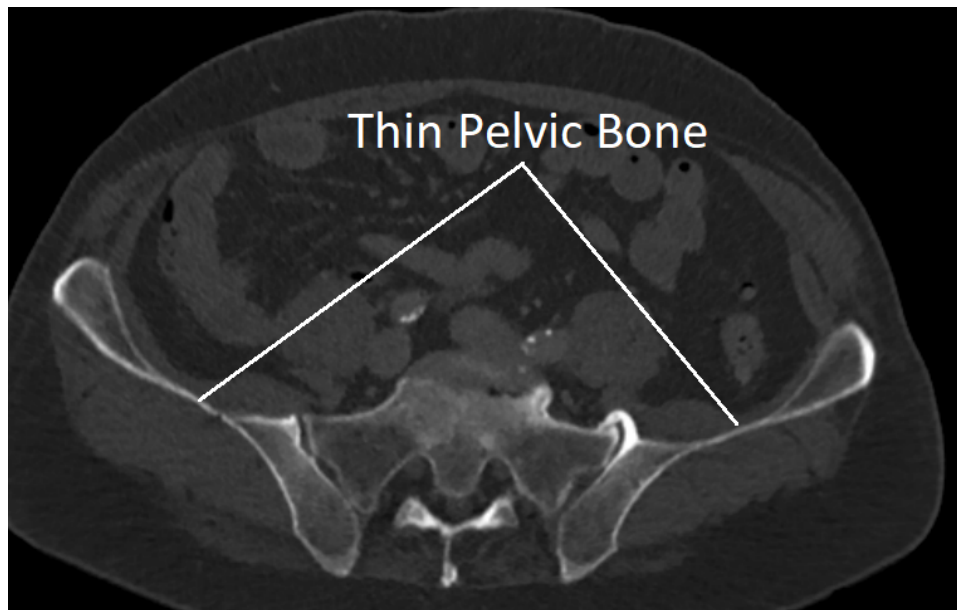


Fig. 2.5 The pelvic bone is extremely thin in places, as can be seen in this CT image. The bone at these locations does not contain any trabecular bone but is entirely cortical bone. Underlying CT data is from the Addenbrooke's Hospital CT Data Collection.

2.2.5 Bone thickness calculation

If the bone is significantly thin (much less than 10.8 mm, the length of line that passes through the trabecular bone), CBM will not generally produce an accurate result. For instance, this occurs in two locations, which include parts of the hip, as well as in most of the ribs. Figure 2.5 shows an example of a very thin area within the hip. In this example, the thin area is entirely cortical bone so there is no trabecular bone density to measure. This case produced poor results in the CBM modelling process. The ribs are also very thin in most places even though they do contain a small amount of trabecular bone. This amount however, is also very small, negatively impacting the CBM modelling process.

If the bone is very thin, there is unlikely to be much, if any, trabecular bone to measure or compare with. These thin regions cannot be used as there is not enough information to accurately determine if the bone at that location is diseased or not. It is therefore necessary to find these regions and discard them so they are not used in the disease detection process.

To find points on the bony segmentation that are very thin, each point's vertex normal was projected into the bone and was intersected with all triangles within the mesh, in order to determine bone depth for that point. If multiple intersections occurred, the closest intersection was used to calculate the depth. Points that contain depths less than 3.0 mm were marked as being too thin, although an exact depth is difficult to determine as this depends on how thick

the cortices are relative to the overall bone thickness. Using this cutoff excludes a number of thin regions, including some part of the ribs, but a significant number of points in the ribs can still be assessed.

2.3 Atlas creation

Using a symmetrical atlas is an essential step in detecting and tracking MBD. In order to investigate symmetrical difference in the skeleton, an atlas was used to both refine the target segmentation as well as to produce a symmetrical mapping.

This is slightly different from the last stage of the hysteresis thresholding method described by Fu et al. [43]. In their method, the atlas mesh was registered to the target segmentation mesh only to refine the result – the deformed atlas became the final segmentation [43]. As mentioned above, I used the atlas to both refine the segmentation and to create a symmetrical mapping between points.

To create the atlas, a healthy CT data set was initially segmented using the threshold feature in the Stradview software [133]. Stradview is an medical 3D data visualisation package that can be used to load most types of DICOM data sets. It has a segmentation feature that can be used to segment bone (and edit the segmentations). It can produce high quality surface meshes from the contour lines that surround the segmentations. It also has a cortical bone mapping feature which can be used to model trabecular bone density from the surface meshes.

Ideally, the atlas would be as similar as possible to a normal data set (i.e. in the right age range and the right sex). For prostate cancer patients, the atlas should be male and roughly over sixty years of age. The atlas I used in this thesis (due to the difficulty of obtaining a full body CT scan from a healthy patient), was of a healthy male in his forties. As the age of this data set is younger than almost all prostate cancer patients, the atlas should match less precisely and therefore work less well than if using an atlas of data from a more similar age. However, older patients will have different degrees of degenerate change, which could also cause less precise matching. Osteopaenia or osteoporosis often exists in the older population as well, which can be an additional potential issue for using an aged atlas.

To improve accuracy, the atlas segmentation was then thoroughly corrected by hand using the Stradview software [133]. This process would not be possible for individual segmentations, since it took over 30 hours to carefully correct this segmentation. The full segmentation of the skeleton was also split into multiple regions by labelling general bone groups (left pelvis, centre pelvis, right pelvis), and bones (left upper femur, right upper femur, individual ribs and individual vertebrae) as seen in Figure 2.6.

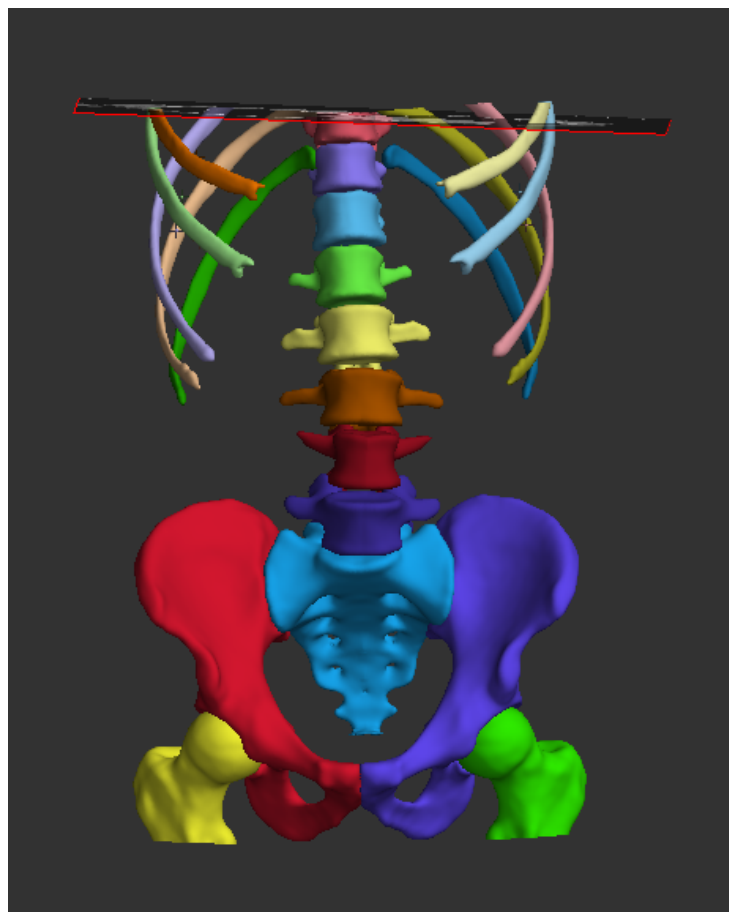


Fig. 2.6 A screenshot taken from Stradview [133] showing one of the atlases. Every point in the polygonal mesh has a perfectly symmetrical point pairing. The colour indicates the individual labelling of bones and bony regions.

For each bony region, a plane was found that best symmetrically split the region in half. In order to find this plane, the Levenberg-Marquardt [90] algorithm was used to optimise plane variables A , B and C in the plane equation $Ax + By + Cz + D = 0$. D was held constant during the optimisation and later used to enforce unit vector constraints required for the plane variables. The Levenberg-Marquardt [90] algorithm requires initial starting parameters, so an initial plane placement was defined by hand. This was found using Stradview [133] for visualisation. After initialisation, all points on one side of the plane (i.e. points that are a positive distance along the plane's normal) were found and projected across the plane as in equation (2.4).

$$P' = P - 2M\hat{N} \quad (2.4)$$

In this equation, P is a selected point, \hat{N} is the symmetric plane's normal unit vector, M is the minimum distance from P to the plane and P' is the projected symmetrical point.

This process was also repeated for points on the opposite side (points a negative distance along the plane's normal). Using these two projections, a complete symmetrical region was formed. The Levenberg-Marquardt [90] algorithm was then used to optimise A , B and C to best reduce the distance between each point in the symmetrical region and its closest point match in the initial region. The error function used was the distance between closest points multiplied by the dot product of the surface normals. This optimisation was iteratively run until near convergence.

Once the plane was optimised, equation (2.4) was applied to every point on one side of the plane to create a perfectly symmetrical atlas. For the set of points that made up triangles that were bisected by the plane of symmetry, a procedure was used to form these points into new triangles (this procedure is demonstrated in Figure 2.7). Then, all sets of points were found, in which two points, O and S , were cut off from a third (R). The symmetric points of O and S were found (O' and S' respectively). Two new triangles were formed, the first being $\triangle OO'S$, and the second being $\triangle SO'S'$. This joined the seam between the selected points and the projected points, creating a full and completely symmetric atlas (Figure 2.8 shows an example of this).

A second atlas could be created by flipping the sign of the symmetrical plane's normal vector and repeating the process, however a second atlas was not needed. This approach was not designed to create an atlas that is exactly true to a real human's bone anatomy, but instead, to create a model that is similar to a human's anatomy but perfectly symmetric.

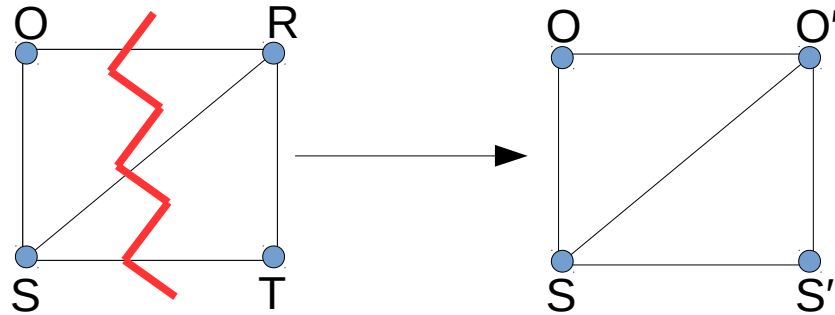


Fig. 2.7 A demonstration of joining the seam between severed triangles when creating the perfectly symmetric atlas. The two triangles $\triangle ORS$ and $\triangle SRT$ on the left are cut (indicated roughly by the red lines) by the plane of symmetry. New triangles $\triangle OO'S$ and $\triangle SO'S'$ are created to join the seam and complete the fully symmetric atlas.



(a) This figure shows an area of the pelvis before the centre seam has been joined together.



(b) This figure shows the same area of the pelvis after the centre seam has been joined together.

Fig. 2.8 An example of joining the centre seam. This process was used to help create a fully symmetrical atlas.

2.4 Registration

Registration is the process of aligning two objects. Objects can vary greatly in type and include 2D images, 3D surfaces, point clouds, landmarks, features, and others. There are three main components of the registration process – a registration (or transformation) model, an objective function and an optimisation method [121]. Of these three, the criteria for the registration model and objective function are the most significant for this thesis, as the optimisation method can be exchanged for a number of workable solutions.

The registration model defines the transformations allowed for the registration process and limits the scope of the registration. There are two types of broad registration models – rigid (which allows for linear transformations and usually a small number of parameters) and deformable (which allows for elastic warpings, and usually includes a much larger set of parameters).

The objective function (or cost function) measures a cost or difference metric between the alignment of the two objects. Any number of difference metrics or matching criteria can be used. Common metrics include using geometric information such as differences in euclidean distance between corresponding points, principal curvatures, surface normal vectors, points of interest, landmarks features and others. Non geometric differences can also be used by measuring the intensity of relationships between two objects. This type of method is often used in multi-modal registration (were the two objects originate from different imaging modalities) [121].

The optimisation method is used to minimise the objective function. Commonly used methods include gradient descent, conjugate gradient, Powell's conjugate directions, Quasi-Newton, Levenberg-Marquardt, and the stochastic gradient descent method [121]. These methods optimise a differentiable objective function (the cost function) over a number of iterative steps. At each iteration, the given method optimises a set of parameters that results in a step in the direction of the minimization. This is repeated until convergence, or until certain criteria are met. These methods vary in step size (which can be constant or variable) and in search direction. The Levenberg-Marquardt algorithm [90] was used in this thesis because the code was readily available. However, any of these methods could be used in place of the Levenberg-Marquardt method with similar results.

2.4.1 Overview of mesh to mesh registration

In order to compare symmetrical bony regions in a CT segmentation for an individual patient, the symmetrical point mapping must be found. Almost all research in this area (finding symmetry) has focused on locating lines or planes of symmetry in 3D point meshes [11].

However, it is difficult to use a plane of symmetry to find 3D point mappings between symmetric regions when looking at the whole skeleton for an individual.

One way to find this symmetrical mapping however, is to register individual symmetric regions (i.e. left femur with right femur) with one of the regions mirrored. After registration, closest pairs of points can be marked as symmetrical point pairings. Segmentations that do not have a corresponding symmetrical segmentation (i.e. the vertebrae) can be registered to a mirrored copy of themselves. This is essentially the process described in Section 2.3. However, this symmetrical point mapping approach requires the target segmentation to be split up and labelled into individual known bones or regions (left femur, right femur, 12th rib left, 12th rib right, etc.). This can be challenging to do accurately using an automatic approach.

The symmetrical point pairing accuracy can be significantly improved by using an atlas-based approach. As the atlas described in Section 2.3 was created to be completely symmetric, it was used to project symmetry into a similarly shaped segmentation. This was accomplished by rigidly registering the atlas to the target segmentation and then by deforming the atlas to best match the target segmentation.

The atlas symmetry can either be projected onto the target segmentation using a closest point approach or by letting the deformed atlas become the segmentation itself, as in Fu et al. [43]. By letting the atlas become the segmentation itself, the segmentation surface is also optimised in the process, which might help improve CBM accuracy. Results of projecting the atlas symmetry onto the target segmentation are described in Section 3.4. Results from using the deformed atlas as the segmentation are described in Section 3.4.12.

2.4.2 Rigid registration

In surface to surface rigid registration, the translation and rotation are found that best minimise the distance between a movable point set P and a static point set Q . For this thesis, P represents the points of the 3D atlas surface while Q represents the points of the 3D target segmentation surface.

This registration technique was first described by Besl and McKay [13] and Zhang [167] and is the basis of all surface to surface point registration methods. In this iterative technique, n steps are taken until there is an error convergence or until some other criterion is reached. In each step, for every point p_i in P , the closest point (usually by Euclidean distance) q_i in Q is found. Then, the least squares solution to the rigid displacement (a translation and rotation) is applied to the points in P as in equation (2.5), which is the best minimisation of the error [130]. This error is defined as the cost function. Having found (R, T) , p is updated, and this update means that the set of matching points will be different in the next iteration.

In equation (2.5), E is the error to be minimised, R is the matrix representing the rotation and T is vector representing the translation. This optimisation problem is often solved using the quaternion method [41].

$$E(R, T) = \min \sum_i \|q_i - (Rp_i + T)\|^2 \quad (2.5)$$

This method can easily fail to achieve good results if the two point sets are at significantly different scales or if they have significant shape differences. To help mitigate the differences in shape and scale, affine registration can be used as described in [42]. Instead of solving for a rotation as in equation (2.5), an affine transformation is solved for as in equation (2.6). In this equation, A is a 3 by 3 (affine) matrix and T is the translation. This affine matrix allows for rotation, scaling and shearing so that P can better register and match Q .

Registration accuracy can also be improved by using a greater number and a wider array of difference metrics. Instead of only comparing the Euclidean distance between points, surface normals and principal curvatures can be used to better match surfaces [42]. This also helps constrain the affine transformation, which can make it more likely to find convergence.

$$E(A, T) = \min \sum_i \|q_i - (Ap_i + T)\|^2 \quad (2.6)$$

2.4.3 Deformable registration

Affine registration can still fail to produce registrations that include significant differences in shape between the two point sets. This is because the affine registration is often not flexible enough to allow for a complete deformation of the points in P to match the points in Q . These differences may include the fine surface details that are important in accurately registering one surface to another, though they can also include errors in the surfaces that are not desirable to register. In order to match the fine details, a deformable registration model was used.

The deformable registration model used in this project was required to meet certain criteria, namely that the method has been shown to work on 3D meshes (not solely 2D images) and that the method does not require manual landmarks. Placing manual landmarks can be challenging and time-consuming (depending on the size of the data set). Often this requires manual interaction to accurately place the landmarks, which relies on an expert's knowledge. This is especially true in a medical setting with complex data sets and unknown pathology. Because of this, a deformable method that does not use manual landmarks was necessary.

There are a number of deformable models that allow for a greater degree of deformation to occur. The primary methods include radial basis functions, elastic body splines, free form deformations and locally affine models. Radial basis functions, while beneficial in interpolating between values, has difficulty seeking local transformations and often requires manual landmarks. A popular version of a radial basis function is the sliding semilandmarks algorithm. It was first introduced by Bookstein [15, 16] and extended to three dimensions by Gunz [51]. In this approach, matching landmarks are found between the point set P and the point set Q . Semi-landmarks are then roughly placed either manually or semi-automatically between the landmarks on the curves and surfaces [50], which are allowed to slide along these curves and surfaces [51]. The thin plate spline matches the landmarks between P and Q and then, as smoothly as possible, it interpolates the data between these landmarks by minimising the total bending energy of the deformation [50]. Using sliding semilandmarks can produce very accurate deformations if the landmarks are placed accurately. I chose not to use sliding semilandmarks because accurate and automatic deformations were needed in order to have a fully automatic MBD detection and tracking algorithm. Also, the scale at which the deformation process needed to accurately work was large (ribs, spine, pelvis and upper femurs) and the bony shapes greatly varied across the different anatomies. This made using sliding semilandmarks on these data sets challenging. Elastic body splines can also be used to optimise matching surfaces, but also requires manual landmarks. Free form deformations are often used as a type of transformation model, however topological preservation is not guaranteed and requires manual control points. Locally affine models use locally linear deformations in order to parametrize transformation that do not require manual landmarks and are suitable for 3D meshes. In a locally affine transformation, each point p_i in P is allowed to register independently of the other points in P . In each iteration, an affine transformation is applied to p_i by finding the best affine transformation that registers a group of points in P that are within a maximum distance R away from p_i to their closest matches in Q [42]. The points in P are then smoothed and the process repeated until a criterion is met. The locally affine transformation has been used successfully in a number of studies [2, 4, 5, 42, 83, 81].

There are a number of other approaches similar to the locally affine model. In a local rigid transformation, each point p_i in P is allowed to register independently of the other points in P . However, at certain intervals, points with similar deformations are clustered together. These clusters of points are then registered as a single group in subsequent steps [59]. This clustering helps add more rigidity to the deformable registration, which can be important in preserving distances between points. A number of studies have used locally rigid transformations successfully [59, 120, 126]. Additional studies similar to the locally affine transformation,

used a local displacement approach where groups of points with like attributes were forced to displace in similar ways [59]. This local displacement approach has been applied in several other studies [29, 92, 146].

Of these methods, the locally affine transformation, the locally rigid transformation and the local displacement field transformation follow a similar pattern. Individual points are free to deform but are influenced by their neighbouring points' deformation in some way in order to help preserve the overall surface. These methods are completely automatic and have been shown to work well in 3D deformable surface registration. Because of the levels of success a version of the locally affine transformation model was used in this thesis.

In addition to these methods, isometric registration is an important deformable registration concept that can help improve accuracy when registering point clouds to each other under certain conditions. Instead of solely using external properties of the surface for registration, internal properties within the surface are preserved in order to help guide the registration [130]. Isometric registration has been shown to be an especially effective tool when registering surfaces containing anatomical morphology [130]. Baiker et al. [8] and Yip et al. [164] created an articulated skeletal joint model used to limit bony registration to known ranges of joint motion in order to guide human and mouse skeletal registration. Isometric registration creates a much more structured registration that can improve accuracy and can help reduce the dimensionality of the registration problem as the deformations are mostly limited to the joints.

2.4.4 Initial placement in registration

A significant source of error in both rigid registration and deformable registration is often due to poor initial placement of the two point sets. This is because a local minimum can easily be found instead of the optimal global minimum [13].

Many approaches have been used to help find accurate initial placements between the two point sets. One very common method is to use a rigid global alignment for a number of steps to initialise the registration before deformable registration takes place [20, 81]. Other approaches try multiple initial placements, and use the best one for the final initial placement [59]. Prior shape information (i.e. the skeleton) can also be used to help the initial placement of the two shapes [1, 98].

A well established method of guiding initial placement in skeletons is to use a piece-wise hierarchical registration [9, 10, 67–70, 74, 117, 164]. This is a prior-based method, as the skeletal bone structure is generally the same across specimens so the general skeletal shape can be leveraged. Piece-wise hierarchical registration was first introduced when registering mice skeletons [8] and was later applied to humans [164].

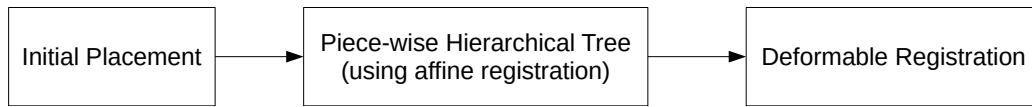


Fig. 2.9 An overview of the registration method.

In this approach, the entire skeleton as a whole is first registered for initial placement. Multiple initial placements are attempted in order to find the best starting initial placement. Then smaller and smaller sub-regions of the skeleton are rigidly registered following a hierarchical anatomical tree of the atlas. Each previous registration acts as the initial placement for the current step, which guides the smaller subsections of the skeleton into a generally correct location before the final rigid registration and deformable registration. This greatly improves accuracy and allows each bony region to move independently of each other but be guided by the skeleton as a whole.

2.4.5 Piece-wise hierarchical registration

For this research, a three part registration technique was used (this can be seen in Figure 2.9). First, the initial placement of the atlas was calculated by using the centroids of the target mesh and the atlas for a rough global positioning estimate. All 103 CT scans within my data-set (described in Section 3.2.1) were oriented the same way with the x-axis, y-axis and z-axis corresponding to the transverse, frontal, and longitudinal axes respectively. This simplifies the initial placement problem, however it is possible for CT data sets not to be in this orientation, although most CT data sets do contain a patient-derived coordinate system.

As the orientations were roughly the same in my case, the atlas was initially placed by lining up centroids between the target and atlas. Then, the atlas was moved by a positive or negative z-offset relative to the centroid. At each z-depth, for each point in the atlas, the closest point (by Euclidean distance) in the target was found. These distances were summed to determine the total distance error. After trying all z-offsets, the position with the smallest error was chosen as the final initial position. Figure 2.10 and Figure 2.11 show two examples of initial placement.

This initial positioning was especially needed when the number of z-slices was significantly different as it was in the case of the healthy control data-set. Often, these CT scans did not include the pelvis and femurs making them much shorter. If the CT scans had differed significantly in orientation, I would have used a principal component analysis to help align principal vectors for better alignment.

After initial placement, a piece-wise hierarchical affine registration technique was used to guide individual pieces of the atlas into relatively correct spatial locations. The euclidean

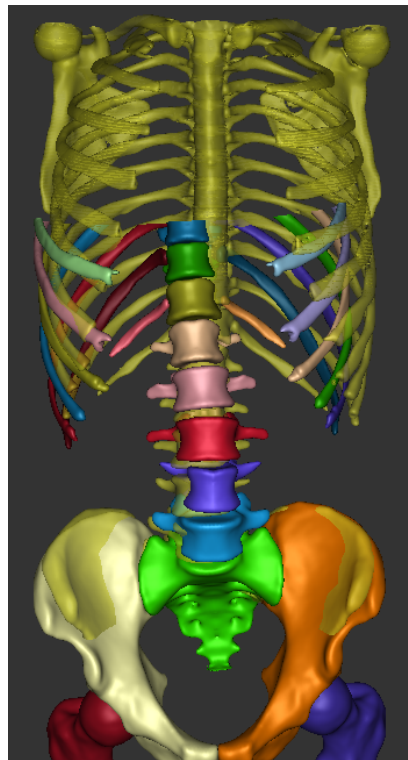


Fig. 2.10 This image is an example of typical initial placement where the atlas is multi-coloured and the target segmentation is coloured yellow. Initial placement in registration is very important as a poor initial placement can lead to inaccurate registration.

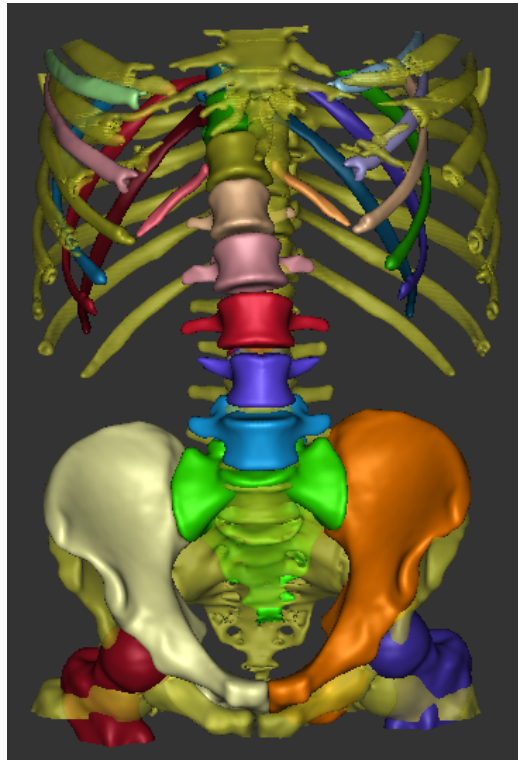


Fig. 2.11 This image shows another example of a typical initial placement.

distance between each point in the atlas and its closest point in the target mesh was used as the cost function. A diagram of the hierarchical tree can be seen in Figure 2.12. At each node in the tree, each bone group was registered together for a number of iterative steps (until only a very small amount of improvement is made in each step) using affine registration. Each previous node and step guides the current registration into place. The figures in Figure 2.13 show an example of piece-wise hierarchical registration, starting from initial placement.

After this, a modified version of the locally rigid transformation was used to deform the final registration so that it closely matched the target segmentation. Instead of starting with individual points, the deformation process started with full sets of the bony regions. These regions were split in half by first finding the points farthest away from each other in the mesh. By treating the points within the mesh as nodes and their triangle relationships as a connected graph, two simultaneous breadth first searches were expanded starting at these initial points until all nodes were claimed by either search. Each of these two regions was registered using affine registration to best align their point sets to Q . The regions were then smoothed using a Gaussian kernel in order to prevent hard boundaries from forming between the patches. This approach helped preserve some of the structure gained from clustering due to the locally rigid transformation. This is useful in covering holes in the target segmentation

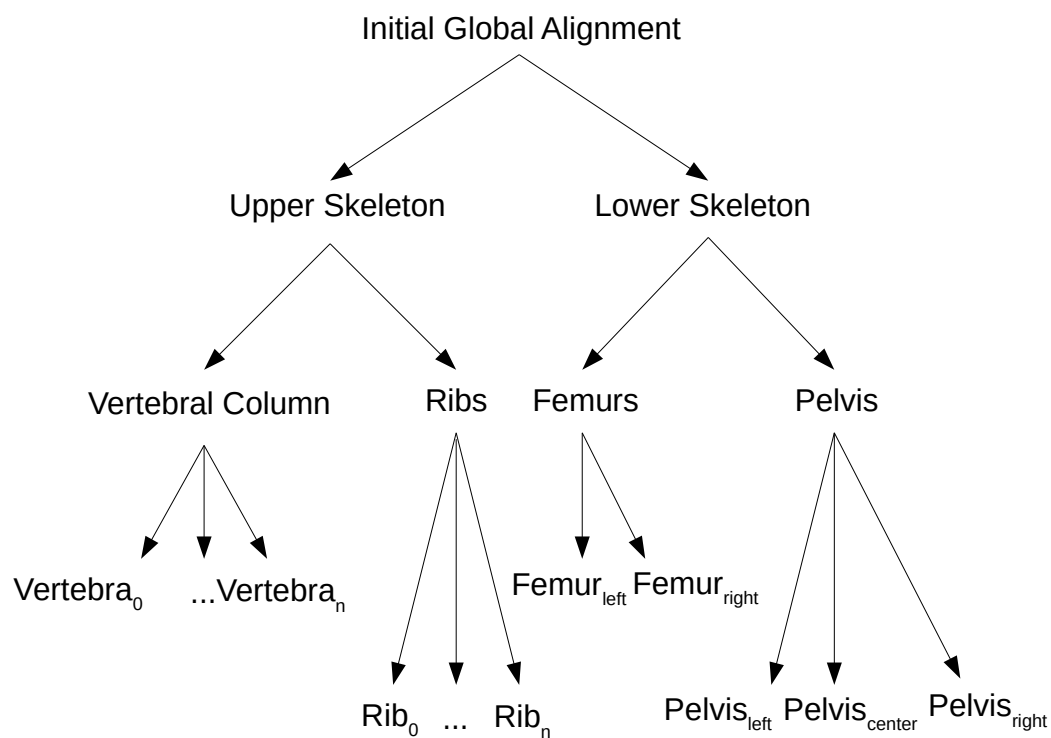
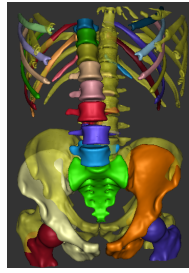
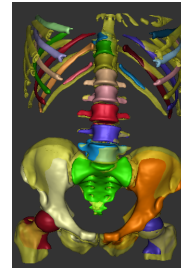


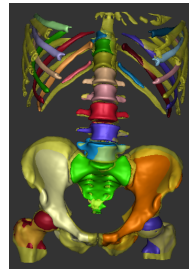
Fig. 2.12 A diagram of the hierarchical registration. Initial registration first occurs globally with the entire skeleton. Progressively smaller regions are then registered with each previous step acting as the initial placement for the current step. Finally, each bone is deformed to match the segmentation.



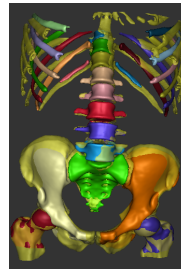
(a) After 0 iterations – the initial placement of the two meshes.



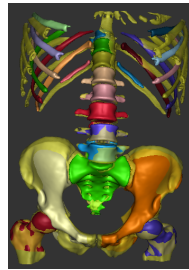
(b) After 25 iterations – registration of the entire skeleton (global alignment).



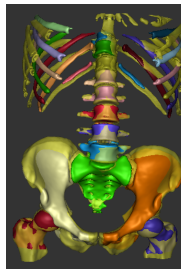
(c) After 50 iterations – registration of the lower skeleton (pelvis and femurs).



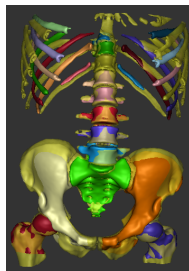
(d) After 75 iterations – registration of the upper skeleton (spine and ribs).



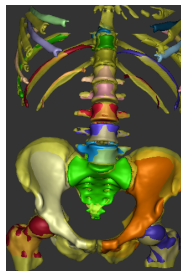
(e) After 100 iterations – registration of the pelvis and then registration of the individual bones within the pelvis.



(f) After 150 iterations – registration of the left femur and then registration of the right femur.



(g) After 200 iterations – registration of the spine, then of the individual vertebrae and then of the individual ribs.



(h) After 300 iterations – the final registration result before the deformable registration process.

Fig. 2.13 An example showing a typical piece-wise hierarchical affine registration process of registering the atlas mesh to the target mesh.

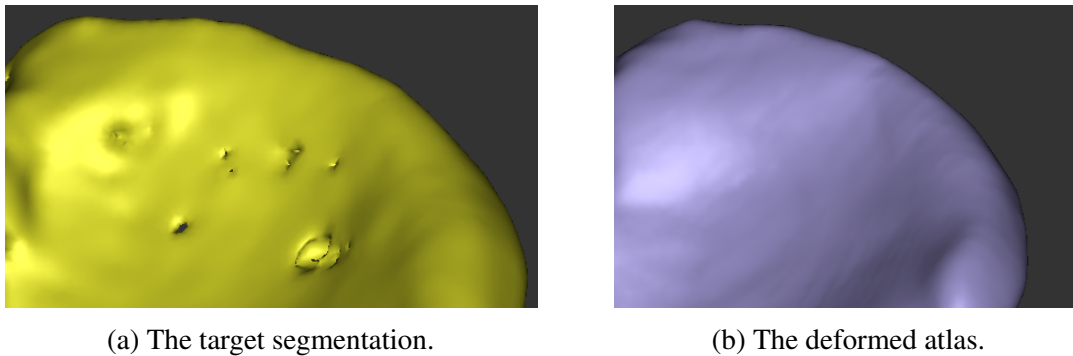


Fig. 2.14 An example of comparing a target segmentation to a deformed atlas segmentation. The target segmentation often contains a number of holes as seen in Figure 2.14a. This figure shows a typical example of the target segmentation mesh result of part of the pelvis. In Figure 2.14b, the deformed atlas was used to cover the holes.

as seen in Figure 2.14. Again, the difference between each point in the atlas and its closest match in the target segmentation was used as the cost function.

The splitting and registration process was repeated for a number of iterations with each iteration creating smaller and smaller patches to deform in order to match finer details. This process continued until little improvement (the reduction of the distance error) was made for a step. Each previous step helped guide the current step into place by providing finer and finer placement. Once the patches were small enough (only a few points), this deformation process behaved very similarly to the locally affine registration. Each point was free to deform but was dependant on the radius of points around it (only in this method they must be connected by the graph). This helped capture the fine details of the registration. Figures 2.15, 2.16 and 2.17 show an example of the deformable registration starting from the last step of the piece-wise registration.

It was also found that optimising both surface normal difference and Euclidean point distance was needed in order to produce accurate registration results. The difference between surface normal vectors of the two points was compared by first normalising the vectors and then by finding the dot product between them. Results from this calculation ranged from -1.0 to 1.0 where a greater value indicated more similar surface normal directions. Points were only used in the registration if their surface normals matched significantly well (for each point on the atlas, its paired point was set to be the closest matching point in the target mesh, where the dot product between the surface normals was greater than 0.5). This threshold was found by trying a selection of values from 0.0 to 1.0 (0° to 90° difference between vectors), although a range of nearby values (to 0.5) produced similar results. Using a higher value will enforce greater surface shape matching, while using a lower value will weight minimising Euclidean point distance to a greater extent.

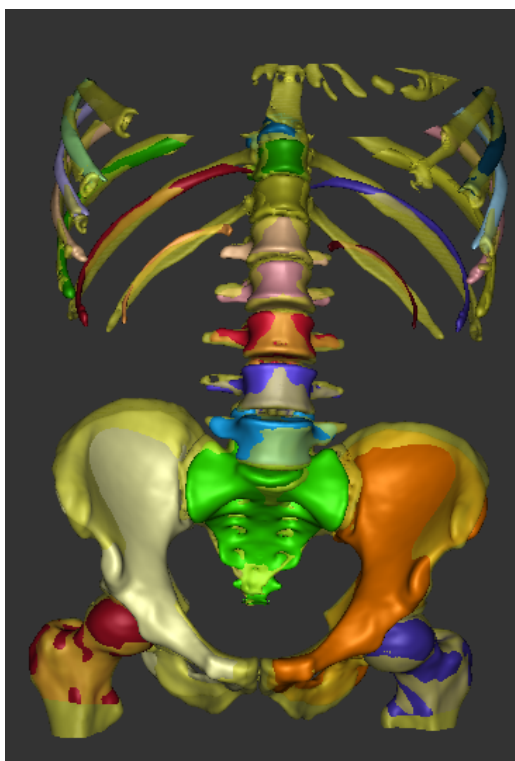


Fig. 2.15 The results of the piece-wise registration before the deformable registration has started.

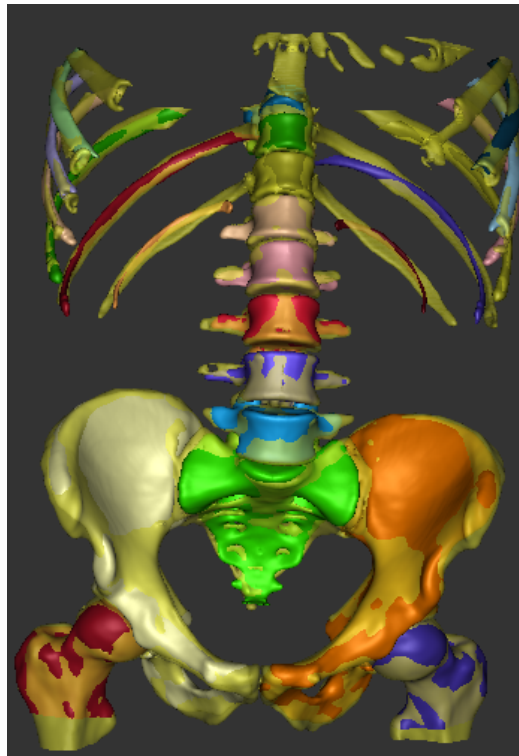


Fig. 2.16 The results of the piece-wise registration after 20 iterations of deformable registration.

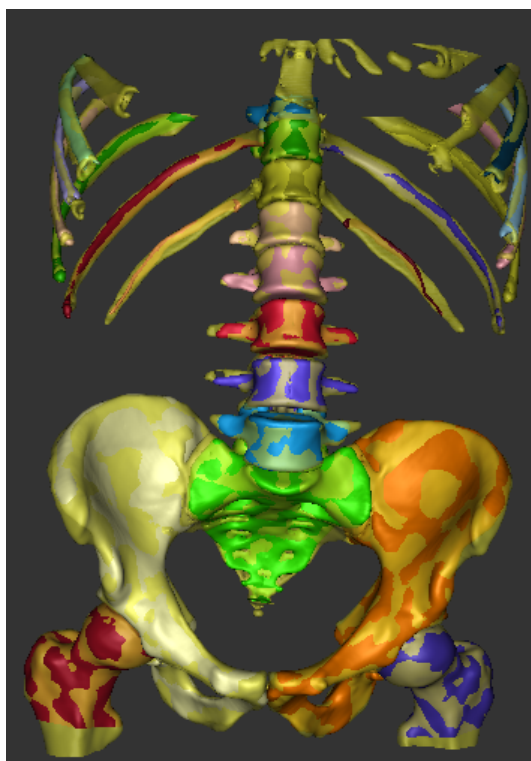


Fig. 2.17 The results of the piece-wise registration after 40 iterations of deformable registration.

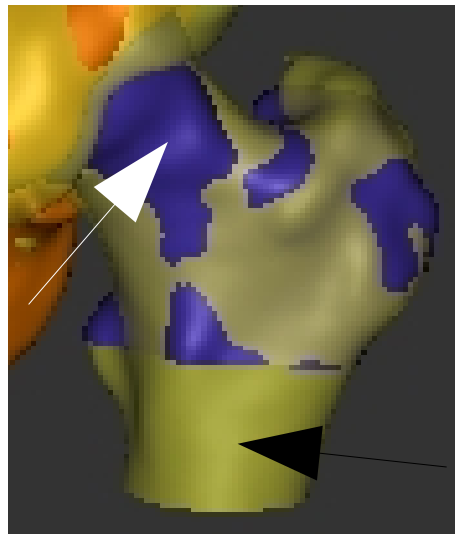


Fig. 2.18 In this image, the best affine registration is found, so the atlas (shown in purple), can best match the target (shown in yellow). The points in the atlas indicated by the white arrow are significantly far away from areas of the target mesh in comparison to the other points in the atlas. One such area is indicated by the black arrow. These points could easily pull the atlas in the wrong direction and so are ignored when computing the registration.

It was also found that as the piece-wise hierarchical model guided each bone into place, its registration was limited similarly to that of an isometric registration articulated joint model, thus using an actual articulated joint model was not necessary. Similarly, matching principal curvatures was not needed when finding closest points, as the piece-wise hierarchical model again guided the registration into the correct locations without examining principal curvatures. However, these two techniques could be used if a more accurate registration was needed.

Additionally, at each step of the registration, the mean Euclidean distance was calculated. If a point was farther than two standard deviations away from the mean, it was considered an outlier and was not used to find the affine transformation. This allowed for points that did not have a good point match to not influence the registration. This was important because it allowed bones of different lengths to register correctly, as seen in Figure 2.18. However, this created other problems as seen in Figure 4.6, for instance, because the corner of the pelvis was prevented from matching as the sides were already close. An example of the final registration after initial placement, piece-wise hierarchical affine registration and deformable registration can be seen in Figure 2.19.

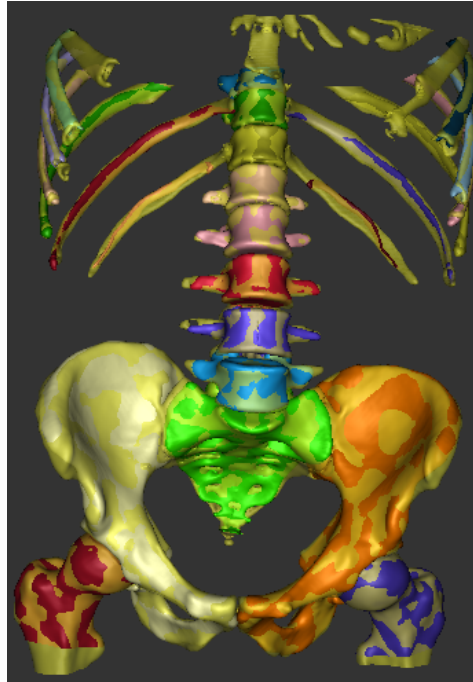


Fig. 2.19 A typical final registration – each region of the deformed atlas is coloured in a different colour while the target segmentation is in yellow.

2.4.6 Improving the hierarchical registration

It was found that even when using a piece-wise hierarchical approach, the registration of the spine still contained a significant number of errors, as the vertebrae were similar in shape and easily moved into the wrong location during registration. If a vertebra was initially placed at too high a z-value, it often would incorrectly slide up and register to the wrong vertebra. Similarly, if the vertebra was initially placed at too low a z-value, it often slid incorrectly down.

For simplicity in this thesis, individual vertebrae are labelled from *Vertebrae*₀, which is the lowest vertebra closest to the pelvis, to *Vertebrae*₈, the vertebra farthest from the pelvis. The mapping of the proper anatomical terms for these vertebrae are listed in Table 2.2.

A number of studies exist, such as Chen et al. [24], Glocker et al. [48, 49], Kelm et al. [66], Klinder et al. [72], Suzani et al. [127] and Yang et al. [159] that have solved for vertebra centroids (vertebra localisation) within CT data sets. This can be used as initial placement for atlas vertebrae. However, due to variable sizes between the target segmentation vertebra and the atlas vertebra, even when having correct vertebra locations, registration can still easily lead to a vertebrae registering above or below the correct vertebra. This is because the top surface of an atlas vertebra often registers with the bottom surface of a target vertebra located a place above its intended registration target vertebra. This occurs when errors in the

Table 2.2 The numerical ordering (for clarity) of the vertebrae used in the registration.

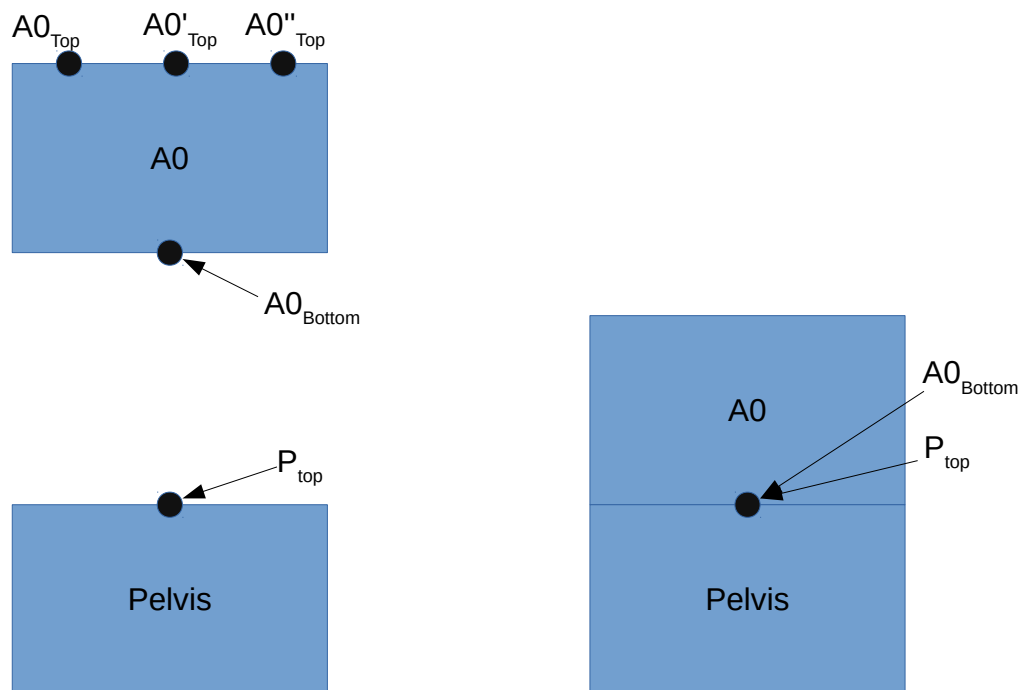
Anatomical label	Project label
L5	<i>Vertebra₀</i>
L4	<i>Vertebra₁</i>
L3	<i>Vertebra₂</i>
L2	<i>Vertebra₃</i>
L1	<i>Vertebra₄</i>
T12	<i>Vertebra₅</i>
T11	<i>Vertebra₆</i>
T10	<i>Vertebra₇</i>

segmentation cause two or more vertebrae to merge into a single body. When this happens, target points and target normals do not always exist on either the top or bottom of individual vertebrae. As registration cannot match onto these missing points, the atlas vertebra can easily slide up and down the merged vertebral column.

After searching the literature, it was found that there has been little written on vertebra to vertebra 3D point registration. In order to solve this issue, three point markers ($A_{n_{top}}$, $A_{n'_{top}}$ and $A_{n''_{top}}$) were placed (before registration) on the top surface of each vertebra A_n in the atlas using a manual visualisation point selection process. Another point ($A_{n_{bottom}}$), was placed on the bottom centre of each, as well as a point (p_{top}) on the top-centre of the pelvis where the spine starts. Figure 2.20a shows an example of this. These markers were used to set the initial placement locations of each vertebra.

Beginning with the vertebra with the lowest z-value (nearest to the pelvis), the initial atlas vertebra A_0 , was placed by lining up its $A_{0_{bottom}}$ point with the p_{top} point on the pelvis. This can be seen in Figure 2.20b. The pelvis was used as a base, as it had already been guided into the correct location using the piece-wise hierarchical tree registration order. The pelvis had a larger and more unique shape that is easier to register correctly. The A_0 vertebra was then registered to its closest target vertebra T_0 , which was the correct vertebra match. This is shown in Figure 2.21.

Then, the second atlas vertebra A_1 (the vertebra right above A_0), was placed by lining up its $A_{1_{bottom}}$ point with the centroid of $A_{0_{top}}$, $A_{0'_{top}}$ and $A_{0''_{top}}$. This is shown in Figure 2.22. Using the plane defined by $A_{0_{top}}$, $A_{0'_{top}}$ and $A_{0''_{top}}$, all points below this plane were ignored during registration. This forced A_1 to register upwards (as there are no point matches below the plane) into the correct location T_1 . This also made it unlikely that A_1 would register with the target vertebra T_2 above its correct location, as the initial location of A_1 was already low (due to the natural vertebra spacing which had been ignored



(a) Points are initially placed by hand on each vertebra (only one shown here) in the atlas and also on the pelvis to mark the top and bottom surfaces.

(b) In this example shown in simplified 2D, the points $A0_{bottom}$ and P_{top} are used for correct initial alignment between the first vertebra and the pelvis.

Fig. 2.20 An example of registering the lowest vertebra with the top of the pelvis. Using pre-marked points greatly improves vertebra registration.

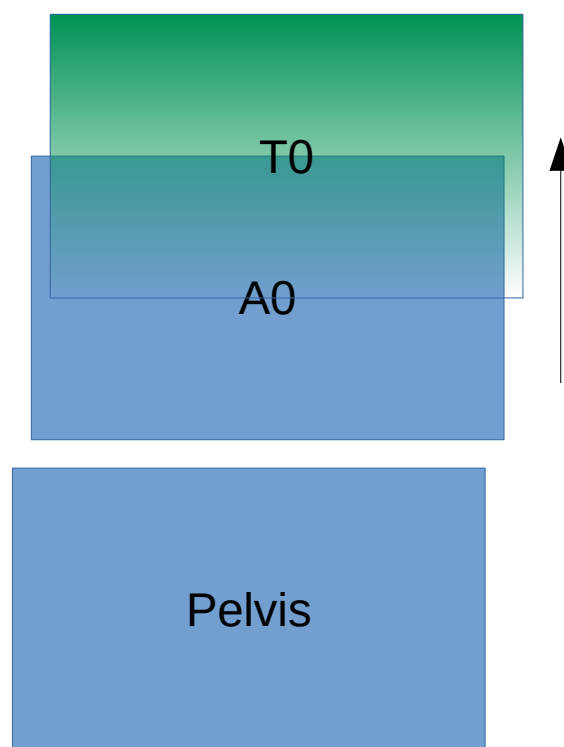


Fig. 2.21 An example showing the registration of an atlas vertebra to a target vertebra. Atlas vertebra *A0* moves upward to register with target vertebra *T0*.

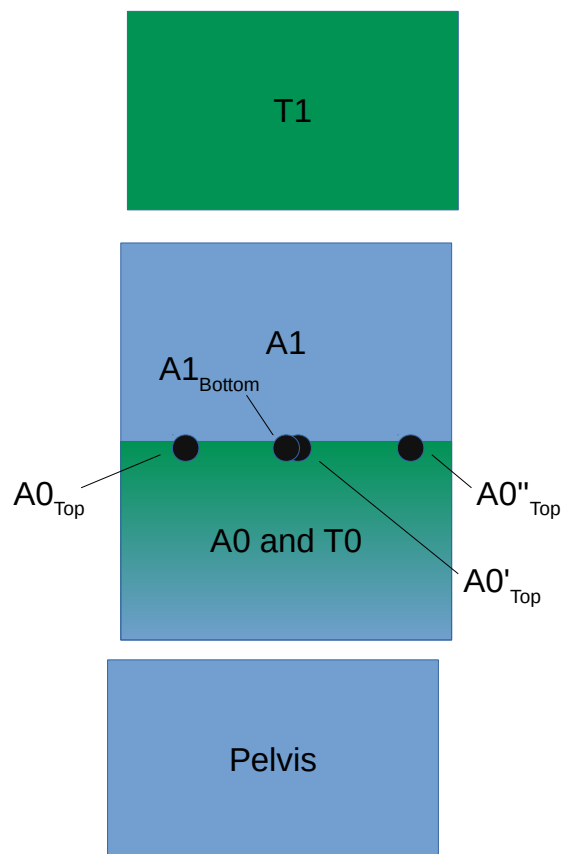


Fig. 2.22 An example of using the top of a previously registered vertebra as the initial starting registration location for the vertebra above it. Atlas vertebra A1 is initially placed by aligning the $A1_{bottom}$ with the centroid of $A0_{top}$, $A0'_{top}$ and $A0''_{top}$.

when placing A_1 right above A_0). This process was repeated for all of the vertebrae. An example of using this technique can be seen in Figure 2.19.

The ribs also often contained significant registration errors as there is great variance between shape, size and placement between scans. The atlas ribs frequently registered to the wrong target rib or incorrectly registered to multiple ribs. Moreover, costal cartilage joins the ribs (not ribs 11 to 12) to the sternum, which can have a variable degree of calcification, especially in the older age group. This may affect the segmentation and therefore the registration. Again, after searching the literature, it appears that very little has been written about 3D rib registration.

To help fix these problems, a technique was used to provide better initial rib placement and to limit each individual rib registration to point groups made of a single target rib. A line was found running through the spine using centre points from the vertebrae with the greatest and smallest z value. This line was used to ignore points on the spine (located within a certain distance of the line) and then split the target segmentation into individual ribs using region growing (breadth first search). A region growing method (Liu et al. [84]) similar to this was used successfully to assign individual ribs unique regions. Once the groups of points that make up the individual target ribs were found, each atlas rib was assigned a target rib point group based on z-ordering and left-right side. Each atlas rib was only allowed to register with its assigned point group. All other points were ignored, forcing the rib to register with its correct rib without interacting with any other rib. This greatly improved registration accuracy.

2.5 Symmetrical disease assessment

Once the segmentation, CBM and registration processes are finished, the trabecular bone density at every point on one side of the segmentation is compared with the other side using the registration mappings. A significant difference in trabecular bone density between the left side and the right side can indicate the presence of MBD.

For this thesis, I implemented a custom viewer written in C++, using the OpenGL [95] library. It can display the results of this method in order to assess the disease visually. It can visualize the polygonal meshes created by Stradview and overlay them with different colour schemes.

In figure 2.23a, the trabecular bone density has been overlaid upon the segmentation mesh. This is not yet showing the symmetrical difference although a user can easily compare the left side of the pelvis to the right side and notice a strong difference in colour. The darker blue indicates a more dense bone while a lighter colour indicates a less dense bone. Different

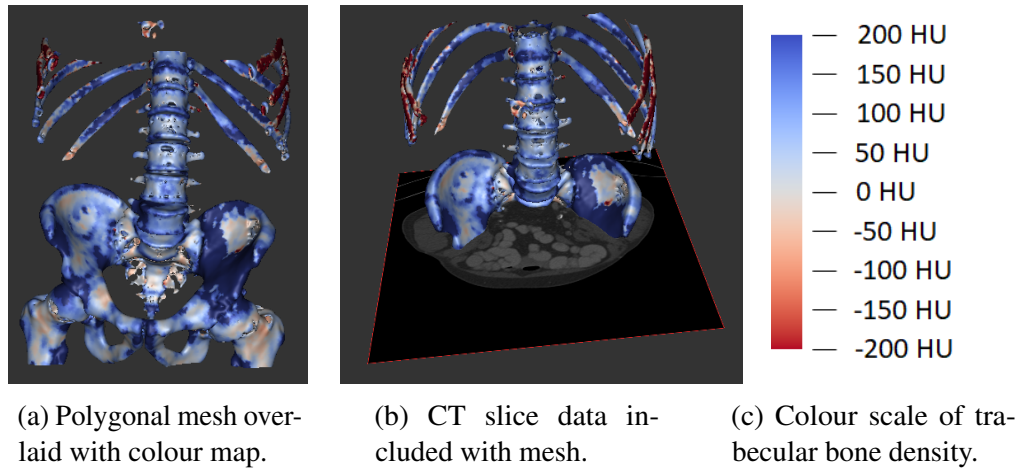


Fig. 2.23 An example of a segmentation result overlaid with a colour map representing the trabecular bone density measurements. A user can easily see the difference in trabecular bone density by comparing the colour of the left side of the pelvis with the right side of the pelvis. This difference can indicate the presence of MBD.

colour projection schemes can also be applied by using a drop-down menu based on user colour preference. This may be an important option for a colour blind user. A user can click and drag in the viewer to rotate the segmentation and to view it from any angle. The user can also zoom in and out to visualise specific regions of the mesh in more detail.

A user can choose any CT slice within the data set and select to either view or to hide the CT image within the viewer window. Figure 2.23b shows an example of when a user has selected to turn on the visualisation of the CT slice image. This allows a user to compare the CT slice images with the trabecular bone density (as it has been projected upon the segmented mesh). A user can also apply an alpha blend to the mesh which blends the background and CT slice image with the segmentation. If the alpha blend is fully applied, the segmentation is completely hidden so only the background and CT slice are visible. An alpha blend can also be applied to the slice image in order to blend it with the segmented mesh.

The CT image in the viewer is distorted due to the projection of the image into the texture map, as the image is viewed using a 3D perspective camera. Because of this, the CT image is also displayed in a separate 2D viewing panel, so a user can view the image directly in 2D. The left and centre panels of the viewer are shown in Figure 2.24.

2.5.1 Smoothing of CBM data

The cortical bone mapping (CBM) process produces an estimate of the parameter variance for each point measurement [136] – this indicates how accurate each trabecular bone den-

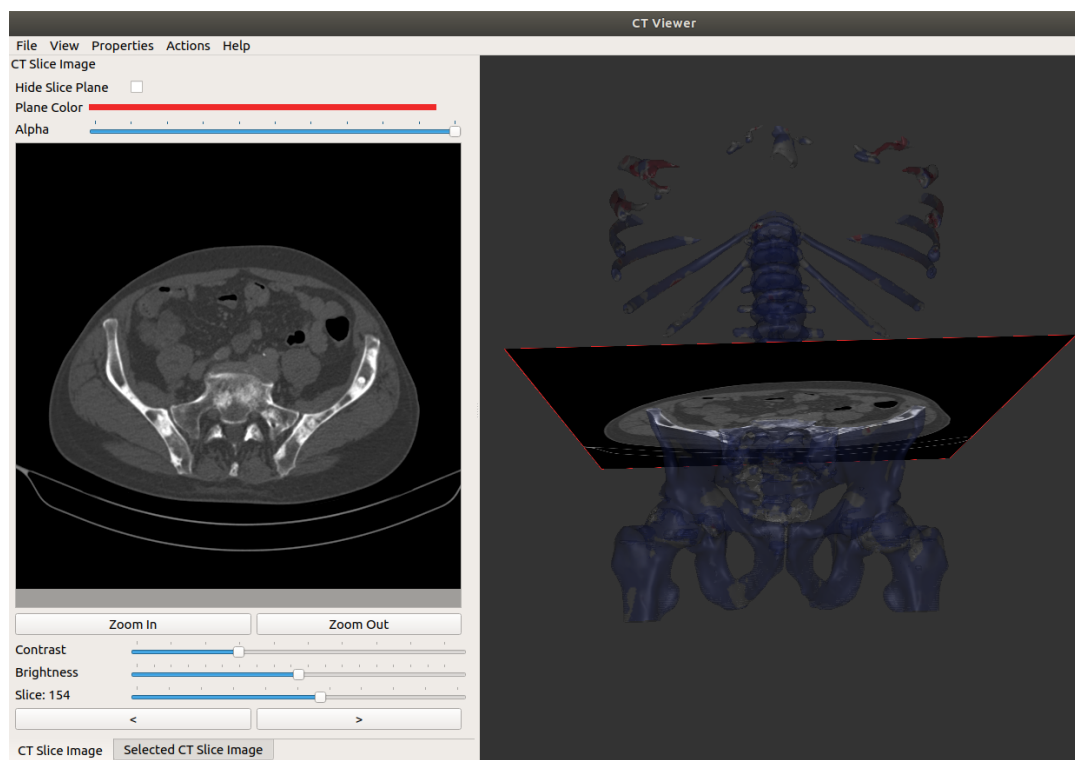


Fig. 2.24 An example of the custom image viewer. The right side of the image shows the main viewer with an alpha blend applied. This blends the polygonal mesh with the background and CT slice (texture mapped) image. The left side shows the CT slice 2D along with some of the user interface controls. This allows a user to specify the display options including the alpha blend, image brightness, image contrast, image zoom and slice image selection.

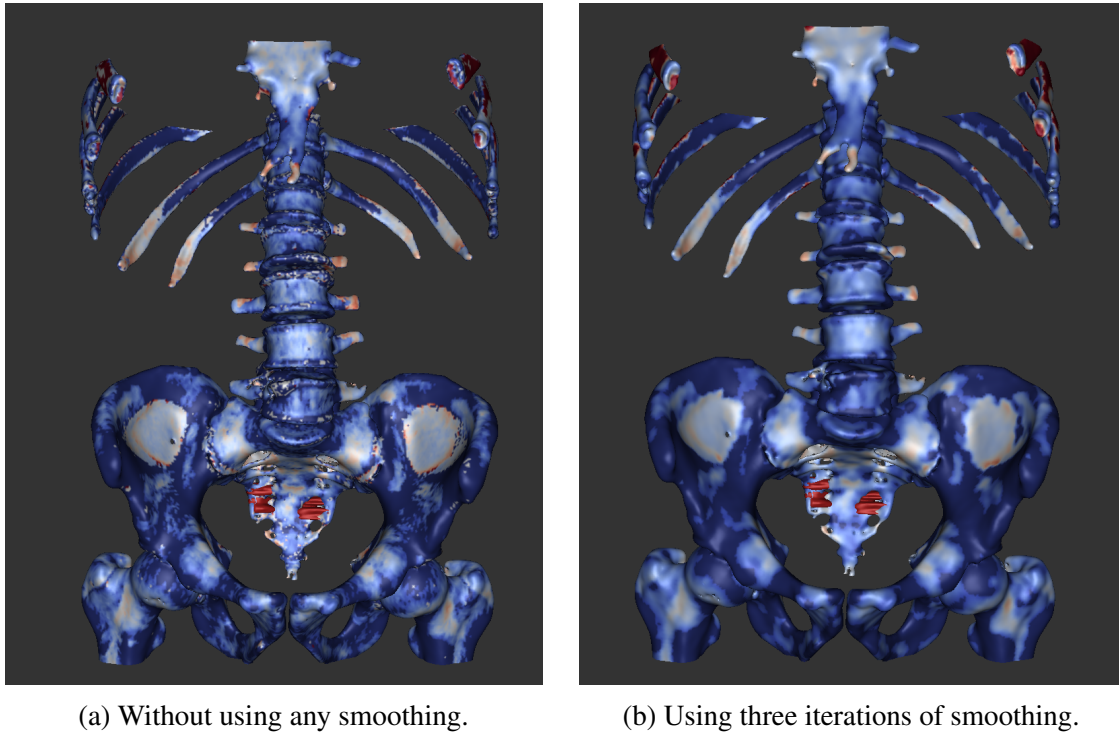


Fig. 2.25 An example of the segmentation where each vertex has been coloured according to its trabecular bone density measurement.

sity measurement was. A smoothing filter weighted by the inverse of these errors can be used to smooth the trabecular bone density measurements across the surface as described by Treece et al. [136]. CBM measurements with lesser error are weighted more than CBM measurements with greater error. This is an important step in improving the accuracy of this algorithm, as it can reduce some of the error in the assessment method.

The filter works by iterating over each point in the mesh and setting the CBM measurement for that point to a weighted (by CBM error) sum of itself and of the neighbouring CBM measurements. This process is repeated a specified n number of times, with the output values of the last iteration used as the input values of the next. Figure 2.25a shows the surface with 0 smoothing iterations while Figure 2.25b and Figure 2.26 show the surface with three and five iterations respectively. The number of smoothing iterations can be set dynamically by the user in the viewer.

2.5.2 Density differences

After the smoothing filter has been applied, the trabecular bone density at each point is subtracted from the trabecular bone density of its symmetrically mapped point (using the

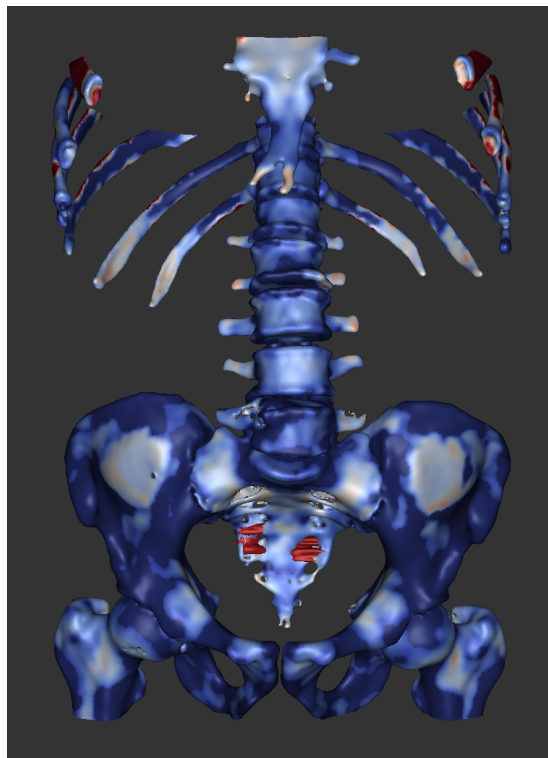


Fig. 2.26 An example of the segmentation where each vertex has been coloured according to its trabecular bone density measurement. This example used five iterations of smoothing.

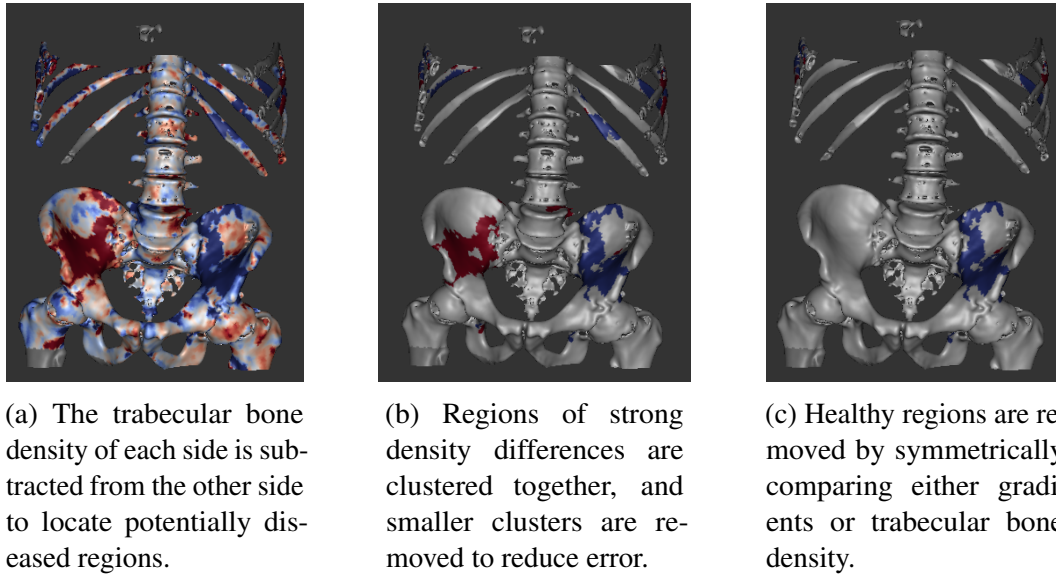


Fig. 2.27 An example of visualising disease by comparing trabecular bone density differences.

symmetrical mappings found in the registration process). This visualisation is shown in Figure 2.27a. In this image, the blue regions indicate a positive difference in trabecular bone density while red regions indicate a negative difference.

A user-selected HU cutoff threshold parameter is then applied to the mesh. This marks a point as healthy if the absolute value of the difference of trabecular bone density for that point is less than the threshold. An example of this is shown in Figure 2.27b.

By treating the polygonal mesh as an undirected graph, where vertices are the nodes, and triangle relationships between these nodes define the paths, connected groups of potentially diseased points are grouped into clusters. A user-selected cluster size parameter is then used to disregard all clusters where the number of points within that cluster is less than the selected cluster parameter. In Section 3.4.4, the number of smoothing iterations, the HU threshold cutoff and the cluster size parameters are optimised.

2.5.3 Diseased side selection

Groups of potentially diseased points could be either healthy or diseased. Healthy regions are still included with diseased regions at this point, as all groups with large differences in trabecular bone density between left and right sides are selected without taking into account which side caused the differences. In prostate cancer most MBD lesions have increased bone density, so the region with the greatest trabecular bone density can be selected as the diseased region. The results of this experiment are described in Section 3.4.10.

However, some cancers cause both sclerotic and lytic lesions to form, which means that both types can exist within a single data set. I wanted to explore a way to automatically distinguish either lesion type. Intuitively, trabecular bone density in healthy tissue is relatively constant, while diseased bone that has increased or decreased bone density will have certain intense changes in trabecular bone density. Therefore, when the gradient of the trabecular bone density of healthy bone is compared to the gradient of trabecular bone density of diseased bone, the gradient of the healthy bone will generally have a smaller value.

By using this principle, I developed a novel method of determining which side is diseased when mixed lesions are present. For each potentially diseased group, the points on the border of each group were located. A diseased point was considered a border point if it was located next to a healthy point. For each of these border points, the gradient of the trabecular density was compared with its symmetrical point's trabecular density gradient. These differences were summed and the group was marked as a lesion if the final sum was greater than 0 (as the group had a greater total change in trabecular bone density than its symmetrical points did).

Very symmetrical regions of actual disease may exist within the bone. However, it is unlikely that their borders will map completely symmetrically to the same place as the regions will be of differing sizes and shapes. This allows for two diseased regions with slightly different borders to both be marked as disease by the assessment method.

In this way, regions can be determined to be either diseased or healthy. In Figure 2.27c, the results of this process is shown, as the healthy side has been hidden from view.

2.5.4 Visualisations of the disease

In the viewer, a user can click on any point on the mesh to find its symmetric point pair. Figure 2.28 shows the results of clicking on the diseased right side. The CT data is automatically re-sliced and shown in Figure 2.29. In this image, the red line shows the surface normal of the point that user has clicked. The blue line shows the surface normal of the selected point's symmetrically mapped point. The green line shows the vector of the line connecting the first and second points. The vectors are automatically calculated and drawn in the viewer on each click. The new, re-slice image plane (shown in Figure 2.29) is defined by adding the two surface normals together to define one axis and by using the green vector as the other. The red text and circle were added by hand in this image for clarity. It is clear that the left side of the pelvis shown in this image is diseased, as it contains a much greater trabecular bone density than the right side does. These re-sliced images can be viewed in a second window (in 2D), in the same way that the original slice images can be, as mentioned above.

The numerical values of the trabecular bone density for both points are also provided in the viewer within another window, along with other information about the points, including

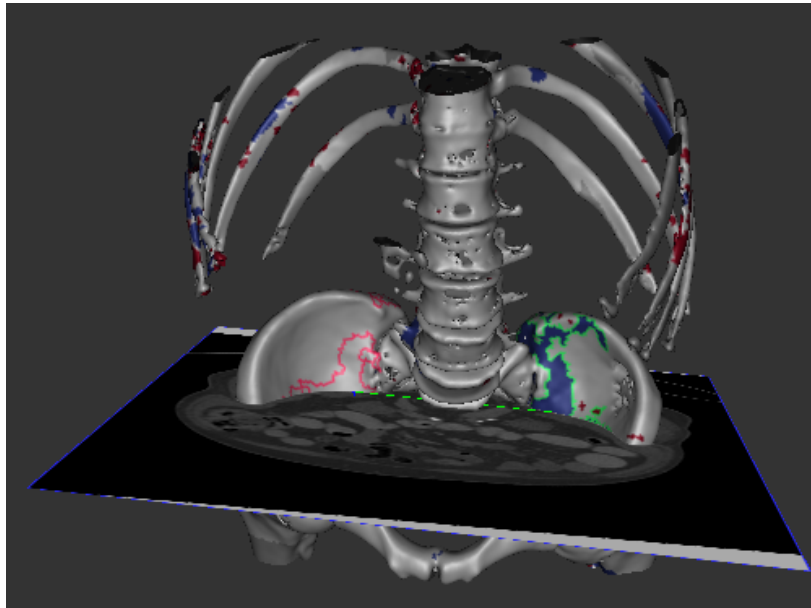


Fig. 2.28 An example of re-slicing the CT data at a user selected location. A user can click on the mesh to calculate a re-sliced CT image at that location. Figure 2.29 shows this resliced CT image without the mesh. The CT data is from the Addenbrooke's Hospital CT Data Collection.

the bone depth, distance between points, other CBM measurements and surface normal vector information.

2.6 Longitudinal disease assessment

A secondary application of this assessment method was developed in order to track disease progression. Instead of using symmetry to compare trabecular bone density within a single case, segmented skeletons of the same patient scanned at different times are registered with each other. Strong differences in trabecular bone density across multiple scans can indicate lesion growth or lesion reduction. The symmetrical atlas was not used in this process as the symmetrical mappings were not needed.

Figure 2.30 shows the trabecular bone density of a baseline case and a follow-up case projected onto the surface of the skeleton. The trabecular bone density is compared between these skeletons by subtracting the density of the registered points. Areas of strong differences are found by using the three parameters (HU cutoff, number of smoothing iterations and cluster size) mentioned in Section 2.5.1 and in Section 2.5.2. This is shown in Figure 2.31 where areas of strong differences in trabecular bone density between the baseline case and the

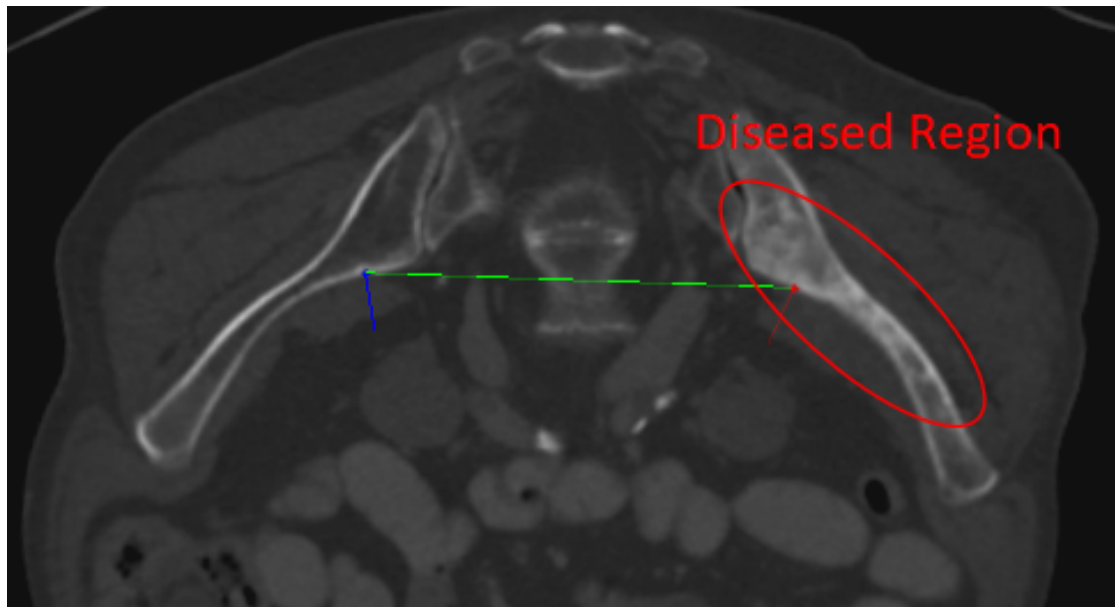
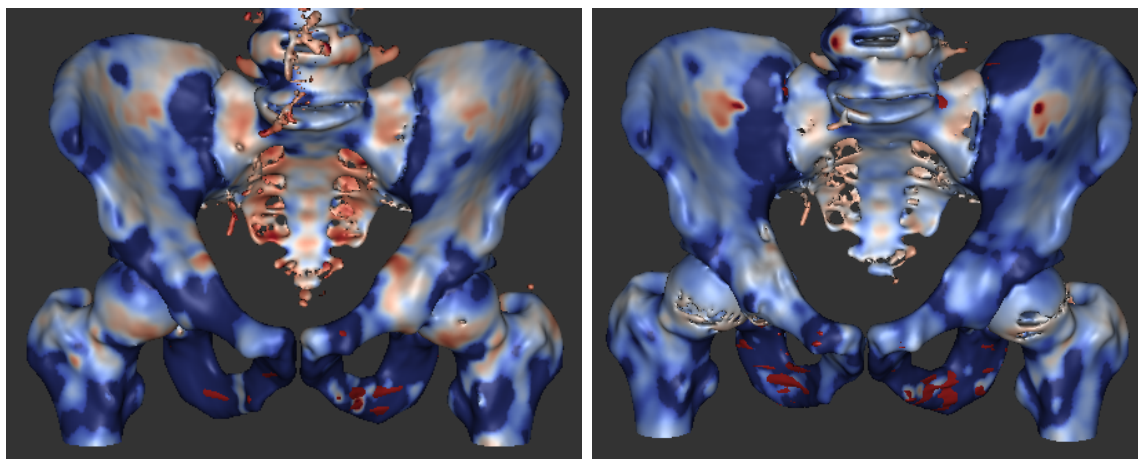


Fig. 2.29 The diseased left side can be seen in the re-sliced CT image. The left side of the pelvis has a much greater trabecular bone density than the right side does. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.



(a) The trabecular bone density of a typical baseline case.

(b) The trabecular bone density of a typical followup case.

Fig. 2.30 In this example, the trabecular bone density from a baseline scan and follow-up CT scan has been projected onto the surface of their respective segmented meshes. By comparing these two images, it can be seen that the disease has progressed, as the trabecular bone density strongly differs in specific regions across these two images. The longitudinal disease assessment method attempts to find these differences.

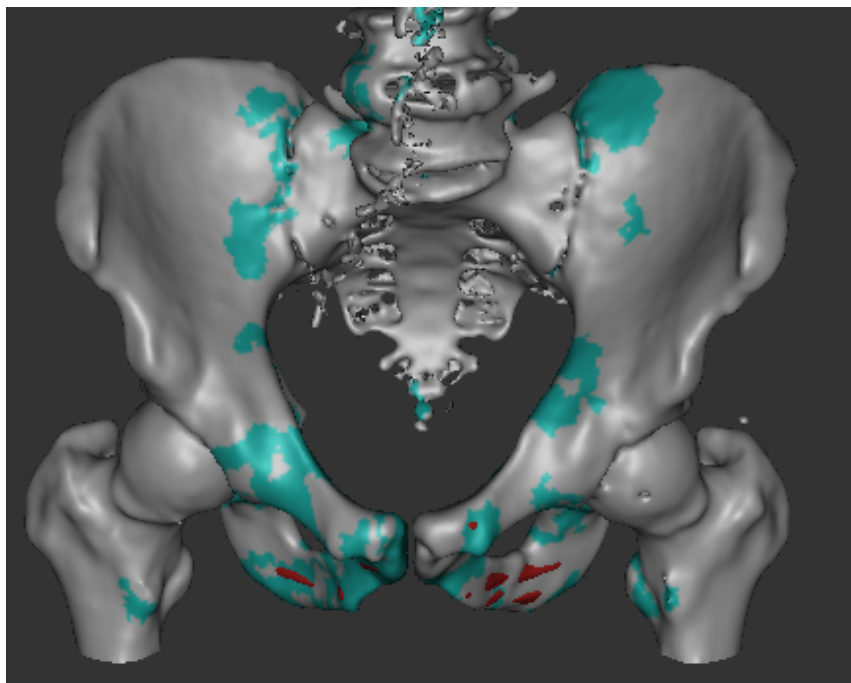


Fig. 2.31 An example showing the disease progression between a baseline case and a follow-up case. The disease progression (in turquoise) has been overlaid over the grey segmentation.

follow-up case have been overlaid in turquoise on top of the segmentation. This highlights the regions where the disease has progressed.

CT HU values may vary across longitudinal scans and across different CT machines, so a calibration step may be needed to improve accuracy when using this method.

Chapter 3

Results

3.1 Introduction

This chapter contains the results of the testing and validation experiments that are used to demonstrate the validity of the detection and visualisation methods presented in Chapter 2. This includes systemic results from both the longitudinal and symmetry disease assessment algorithms, as well as results from individual parts of the system, such as the bone segmentation and registration.

3.2 CT Data

3.2.1 Acquisition

Computed tomography DICOM data was acquired from Addenbrooke's Hospital for use in all the experiments presented below (which is labelled as the Addenbrooke's Hospital CT Data Collection in this thesis). A medical ethics agreement allows for this data to be used for medical research – including for this project. The release of the data used in this project was approved by the Cambridge Urology Translational Research and Clinical Trials (CUTRACT) team [32].

A total of 103 anonymised CT DICOM data sets were obtained predominantly from the DIAMOND database – a research database of patients who have agreed to let their data be used for medical research. Of this total, 68 data sets were acquired from diseased prostate cancer patients and 35 data sets were acquired from healthy patients. In all, the data was scanned from 81 unique patients as 22 patients had both a baseline and a follow-up scan. All patients were male with an age at time of scan ranging from 44 to 89 years old. The average age in the data set was 71.47 years with a standard deviation of 8.51 years.

All diseased data sets were of the full upper body, which included the ribs, spine, pelvis and upper femurs, although some of the data sets from healthy patients did not include the femurs and lower regions of the pelvis. The slice images from these scans had a dimension of 512 by 512 voxels with the spacing between slices 2.0 mm for the diseased data and 2.5 mm for the healthy data. The CT images had an average x and y pixel spacing of 0.776 mm. The average mAs (milliamperere seconds through filament of x-ray tube) for these data sets was 152.857 while the average kVp (kilovoltage peak) was 122.5.

3.2.2 Reference data

In order to have ground truth disease markings to compare with, expert radiologist and secondary supervisor Dr. Barrett marked every diseased region in each 2D image slice of the CT data by outlining each lesion with a contour line. Dr. Barrett is a consultant radiologist, with urology fellowship training and 14 years reporting experience. Dr. Barrett spent an average of 45 minutes per case to create these outlines. Cases with fewer diseased regions took between 25 to 30 minutes, while significantly diseased cases took much longer than an hour. As Dr. Barrett normally spends about 10 minutes assessing a CT scan during a normal radiological CT review, significantly more time and attention was given to marking and creating these outlines than would normally be given to a case. A discussion about the CT data and about the reference data is located in Section 4.2.

Figure 3.1 shows an example of an outline produced by Dr. Barrett that marks the location of a lesion within a CT slice. Figure 3.2 shows another example of an outlined lesion which is much smaller and more difficult to locate. Both of these examples are typical of the lesions and outlines contained within the data.

The 2D contours in each data set were converted into 3D mesh volumes using the regularised marching tetrahedra technique [138]. The Stradview software [133] implementation of this algorithm was used to create these 3D meshes as described in Section 2.2.4. In Figure 3.3, the 3D volumes have been exported from Stradview and imported into a viewer created specifically for this research. These 3D volumes are used as the ground truth disease locations in the experiments described within this thesis.

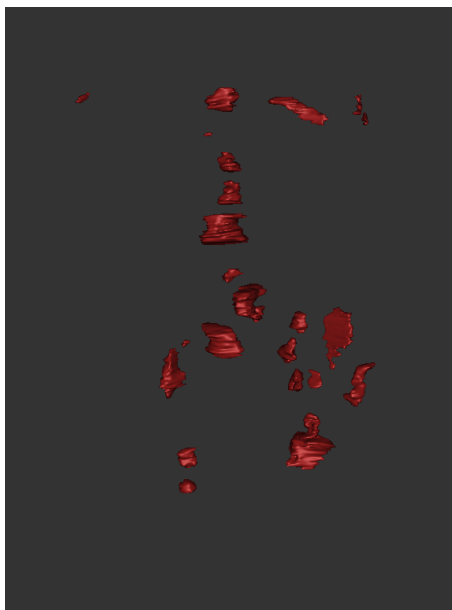
When assessing detection methods, lesion volume size is an important metric, as smaller lesions are most often more difficult to locate than larger lesions. To calculate the internal volume within a triangle mesh (and thus the volume of a lesion), equation (3.1) was used as described by Desbrun et al. [33]. In this equation, V is the total volume, v_i is the i th triangle in the mesh, v_{ij} is the j th vertex within the i th triangle and n is the number of triangles.



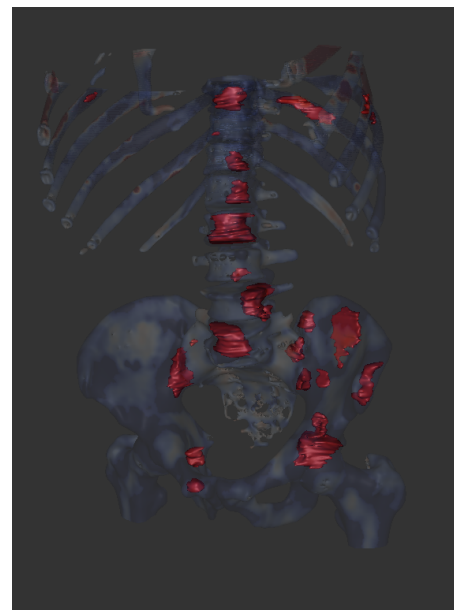
Fig. 3.1 This slice image shows an example of a diseased region outlined by Dr. Barrett in red. Every lesion within each data set was outlined by expert radiologist Dr. Barrett. The lesion is in the left superior pubic ramus bone of the pelvis. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.



Fig. 3.2 Another lesion outlined in red by the radiologist: this one is more challenging to find. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.



(a) An example showing the visualisation of the diseased volumes within a CT data set.



(b) The same diseased volumes overlaid with a faint alpha blend of the segmentation.

Fig. 3.3 Visualisation of diseased 3D volumes created from the radiologist's 2D contour markings. Images are screenshots taken of the visualisation tool which was created for this project.

Table 3.1 The number of lesions binned into volume ranges. The majority of the lesions in this data set are less than 3.0 cm³.

Volume (cm ³) range	Number of Lesions
0 ≤ volume < 3.0	1161
3.0 ≤ volume < 6.0	128
6.0 ≤ volume < 9.0	67
9.0 ≤ volume < 12.0	34
12.0 ≤ volume < 15.0	23
15.0 ≤ volume < 18.0	15
18.0 ≤ volume < 21.0	16
21.0 ≤ volume < 24.0	14
24.0 ≤ volume < 27.0	18
27.0 ≤ volume < 30.0	14
30.0 ≤ volume < 33.0	7
33.0 ≤ volume < 36.0	8
36.0 ≤ volume < 39.0	4
39.0 ≤ volume < 42.0	8
42.0 ≤ volume < 45.0	4
45.0 ≤ volume < 48.0	7
48.0 ≤ volume < 51.0	5
51.0 ≤ volume < 54.0	2
54.0 ≤ volume < 57.0	0
57.0 ≤ volume < 60.0	0
60.0 ≤ volume	33

$$V = \sum_{i=1}^{i=n} \frac{1}{6} (v_{i1} \times v_{i2}) \cdot v_{i3} \quad (3.1)$$

The diseased data sets contained a total of 1,568 distinct lesions (3D mesh volumes). The volume of these polygonal meshes ranged from 0.00296 mm³ to 256.765 cm³ with an average volume being 7.55 cm³. In Table 3.1, the lesions have been binned by volume. The majority of the lesions have a volume less than 3.00 cm³, although a significant number of the lesions have a volume much greater than this. Table 3.2 shows the lesions smaller than 3.00 cm³ binned by volume. Most of the lesions in this data set are located in the spine, pelvis and upper femurs as shown in Figure 3.4. According to Dr. Barrett, the majority of the lesions in this data set are sclerotic.

Table 3.2 The number of lesions binned into volume ranges less than 3.0 cm³.

Volume (cm ³) range	Number of Lesions
$0 \leq \text{volume} < 0.1$	262
$0.1 \leq \text{volume} < 0.2$	191
$0.2 \leq \text{volume} < 0.3$	130
$0.3 \leq \text{volume} < 0.4$	73
$0.4 \leq \text{volume} < 0.5$	63
$0.5 \leq \text{volume} < 0.6$	61
$0.6 \leq \text{volume} < 0.7$	42
$0.7 \leq \text{volume} < 0.8$	44
$0.8 \leq \text{volume} < 0.9$	29
$0.9 \leq \text{volume} < 1.0$	37
$1.0 \leq \text{volume} < 1.1$	16
$1.1 \leq \text{volume} < 1.2$	23
$1.2 \leq \text{volume} < 1.3$	23
$1.3 \leq \text{volume} < 1.4$	10
$1.4 \leq \text{volume} < 1.5$	22
$1.5 \leq \text{volume} < 1.6$	12
$1.6 \leq \text{volume} < 1.7$	13
$1.7 \leq \text{volume} < 1.8$	9
$1.8 \leq \text{volume} < 1.9$	13
$1.9 \leq \text{volume} < 2.0$	13
$2.0 \leq \text{volume} < 2.1$	9
$2.1 \leq \text{volume} < 2.2$	9
$2.2 \leq \text{volume} < 2.3$	15
$2.3 \leq \text{volume} < 2.4$	10
$2.4 \leq \text{volume} < 2.5$	4
$2.5 \leq \text{volume} < 2.6$	6
$2.6 \leq \text{volume} < 2.7$	3
$2.7 \leq \text{volume} < 2.8$	4
$2.8 \leq \text{volume} < 2.9$	5
$2.9 \leq \text{volume} < 3.0$	10

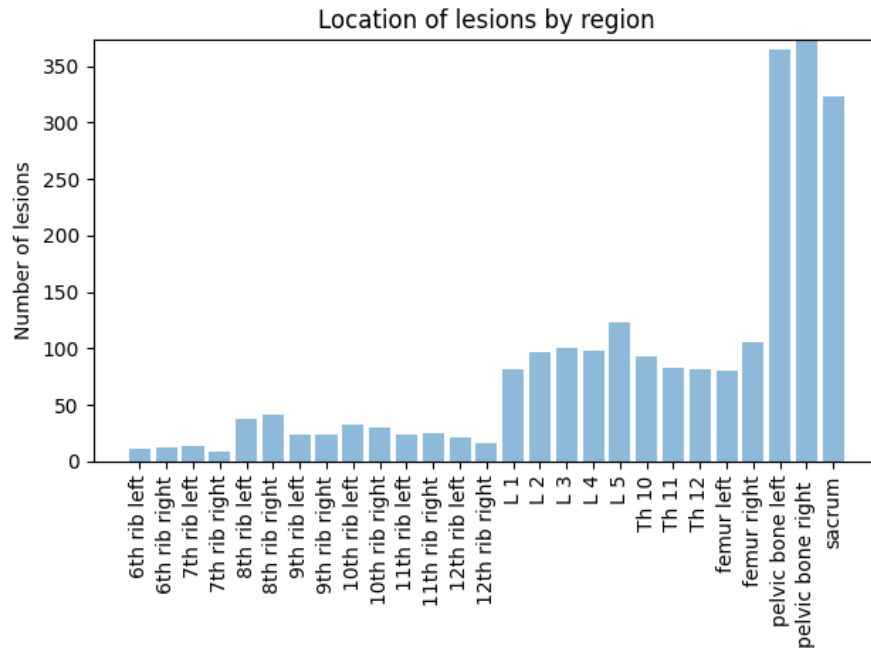


Fig. 3.4 Location of lesions by region (rib, femur, etc.).

3.2.3 Disease shadow size

To project the 3D diseased volumes onto the 3D segmentations of the bone, a ray-triangle intersection method was used. Starting from each vertex within every segmentation mesh, the vertex normal (found by averaging the normal vectors of the triangles that contained the vertex) was projected into the bone and tested for intersection with the 3D lesions associated with its segmentation.

For these experiments, the total line length used in the cortical bone mapping process was set to 18 mm. Of this line, 7.2 mm lies outside the cortex, while 10.8 mm passes through the cortex and into the trabecular bone. If the vertex normal ray intersected a lesion within the 10.8 mm bound, the vertex was labelled as a ground truth diseased point on the segmentation. As shown by the experiments in Section 3.3.2, it is not worth using a line length greater than 18 mm, as all but three lesions are located within 10.8 mm of the surface. These diseased regions on the segmentation surface are called disease shadows. Figure 3.5 shows a typical example of the disease shadows produced by using this process.

The morphology of the bone will constrain both the shape of the disease shadow as well as the maximum size that a diseased volume can grow. Because of this, a single lesion may produce many groups of disease shadows depending on the shape of the bone anatomy. In

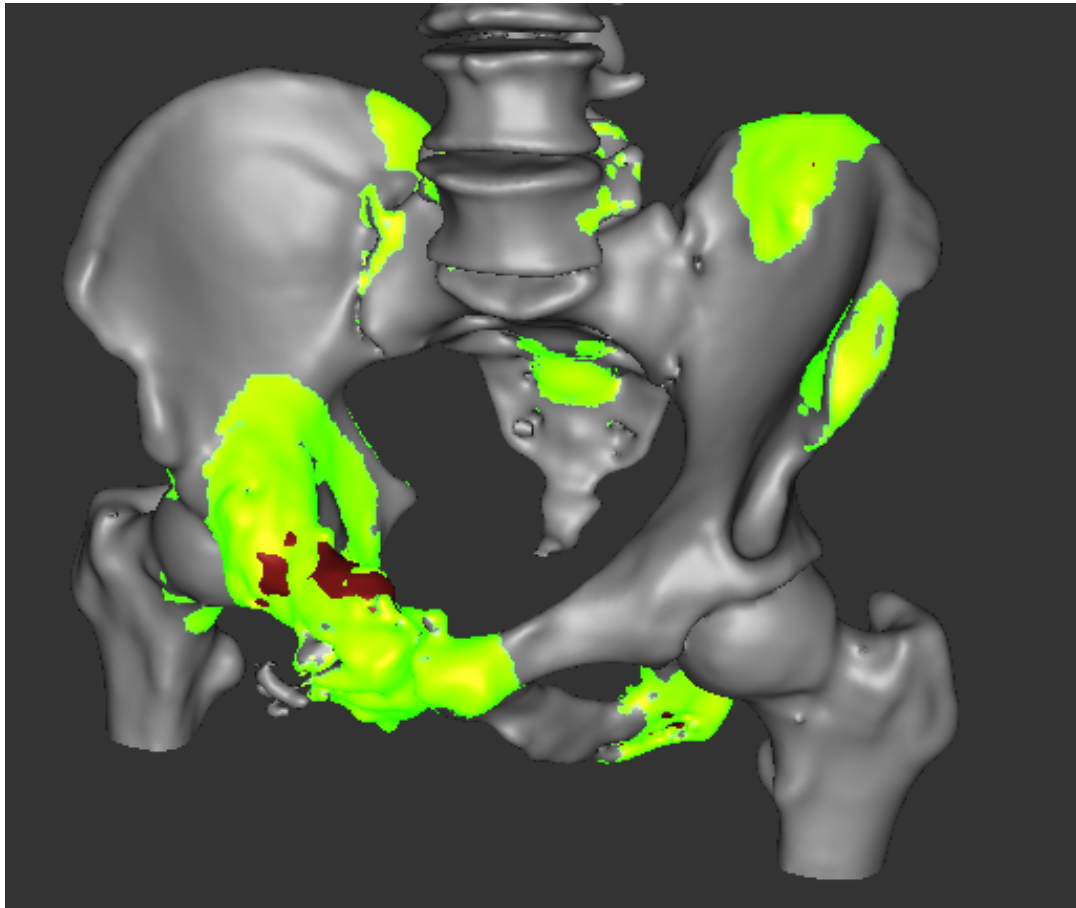


Fig. 3.5 The diseased shadows on the segmentation mesh. The diseased vertices were found by intersecting each vertex normal on the segmentation with the triangles in the 3D lesions less than 10.8 mm away. The disease vertices are highlighted in fluorescent green (the red regions are protruding diseased regions). These projected diseased regions are called disease shadows.

this thesis, the disease shadow size for a lesion is defined as the sum total of the number of the disease shadow points produced by that lesion.

As the volume increases, so does the average number of points in the disease shadow. This relationship can be roughly modelled using a power of $2/3$ as the number of points in a disease shadow is a measure of surface area in cm^2 while the volume is measured in cm^3 . The equation for this relationship was solved for using a least squares approximation method and found to be equation (3.2).

$$\text{disease shadow size} = 288.955 \times \text{volume}^{2/3} \quad (3.2)$$

It is important to note that all of the meshes were created using the same mesh resolution setting (high) in Stradview. This caused the points within each mesh to have the same areal density. Without this, equation (3.2) could not be applied to the mesh in each data set, as the distance between points would differ.

From equation (3.2), 1 cm^3 of volume corresponds to roughly 289 points which is about 17 points per cm length. This is a spacing of 0.6 mm between each point.

In Figure 3.6, lesion volume has been plotted against disease shadow size (number of points) along with the fit line from equation (3.2). The variance of individual data points away from this fit line is most likely due to the differing shapes due to differing anatomy of the various bony surfaces. A discussion about this is located in Section 4.2.5.

If equation (3.2) is reversed, so that volume is described in terms of disease shadow size, then the disease shadow size can be used to provide an estimate of volume. Equation (3.3) shows this relationship.

$$\text{volume} = 0.000204 \times (\text{disease shadow size})^{3/2} \quad (3.3)$$

As the assessment methods described in this thesis produce disease shadow size, equation 3.3 can provide an estimate for the total lesion volume in a data set. To determine the accuracy of this equation, the disease shadow size from every lesion was passed into equation 3.3 and the output compared to the actual volume size. It was found that there was an absolute percentage difference of 52.2% between the actual volume and the predicted volume.

3.2.4 Converting disease shadow points to surface area

To find the average area per point of a disease shadow, the area of each triangle in multiple segmented meshes was calculated using Heron's formula [3]. As the x , y , z cartesian

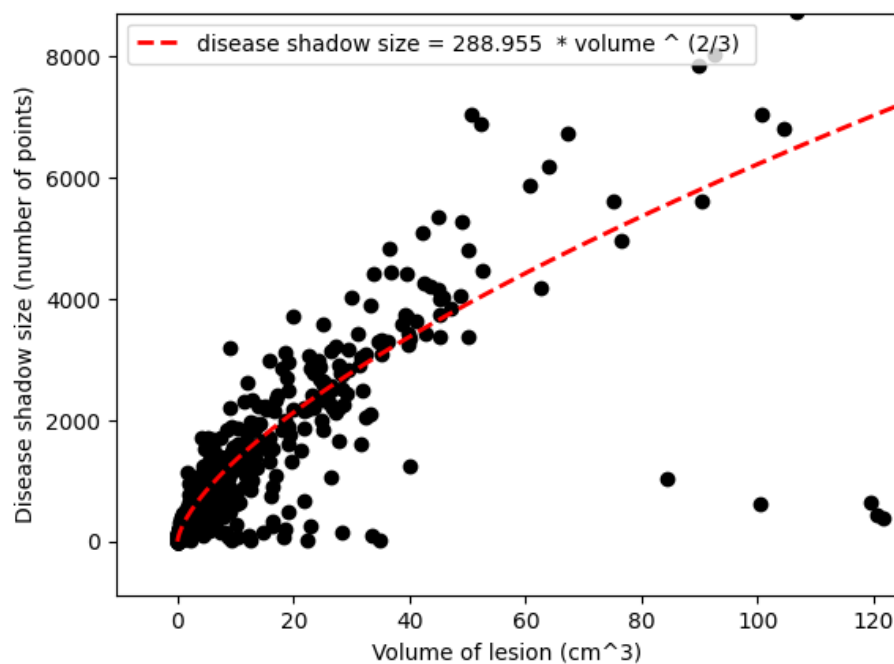


Fig. 3.6 A plot showing the relationship between lesion volume and disease shadow size. As the volume of a lesion increases, so does the disease shadow size projected by that lesion. The fit (shown in red) can be modelled using a power of $2/3$ relationship. This is due to the conversion from volume in cm^3 to number points (which is a measure of area in cm^2). This fit was found using a least squares approximation method. A single 3D lesion can produce many disease shadows so the number of points on the y axis is actually the sum of the number of points in all of the disease shadows produced by that lesion.

coordinates of the points within these triangles correspond directly in mm to their location in real world coordinates, the areas calculated by this formula corresponds directly to the real world area in mm^2 .

To find the average area per point, the sum total of all of the triangle areas was divided by the number of unique vertices (a single vertex can be a point in multiple triangles) in the meshes. 228,622 vertices were used in this calculation, which produced an average area per vertex of 4.090 mm^2 .

When applying this average surface area to a diseased region, the number of points in the disease shadow can be multiplied by 4.090 mm^2 to get a rough estimate of the surface area on the bone. If the number of points in a disease shadow is very small, two points can be removed in this calculation in order to get a slightly more accurate result as at least three points are needed to create a single triangle.

It was also found that the average distance between points was 1.544 mm. This was somewhat different than the rough estimate of 0.6 mm calculated in Section 3.2.3. This difference is likely due to the significant variance found when calculating the disease shadow metrics.

3.3 Evaluation of individual sub-methods

3.3.1 Hysteresis segmentation accuracy

To determine the accuracy of the hysteresis bone segmentation technique as described in Section 2.2.2, six diseased cases representative of the 68 diseased data sets were chosen. The bones in the CT images for these six cases were manually segmented using the Stradview software [133] and carefully checked for accuracy in order to create the ground truth segmentations. It is important to note that in certain bony regions, it is challenging to know what the ground truth segmentation should actually be. An example of a CT image that contains bone that is difficult to segment (even manually) is demonstrated in Figure 2.2.

The CT data for these six cases was processed by the hysteresis segmentation algorithm and the results evaluated against the ground truth segmentations. To determine accuracy, two methods of comparison were used – the dice similarity coefficient (equation 2.1) and the average surface distance (equation 2.2). The DSC measures the correctness of the segmentation using voxel comparison while ASD measures the correctness of the segmentation by evaluating 3D surface distance errors.

The DSC results for six cases (containing 1754 slice images and 459,800,576 voxels) are listed in Table 3.3. The average DSC for these cases was 98.95% This metric is the

Table 3.3 The DSC metric and specificity for six segmentations. Specificity is the number of true positives divided by the (number of true positives plus the number of false positives). The DSC measurement is the ratio of correctly labelled voxels to total number of voxels.

Case	1	2	3	4	5	6
DSC	98.96%	99.14%	99.25%	98.91%	99.03%	98.44%
Specificity	0.92	0.92	0.93	0.90	0.93	0.93

Table 3.4 The average surface distance metric for the six segmentations. The number of invalid points are the number of segmentation points that lie more than two standard deviations away from a ground truth point.

Case	1	2	3	4	5	6
Average error (mm)	2.06	1.67	1.84	1.90	2.75	2.58
Standard deviation (mm)	4.18	4.29	3.26	3.15	4.66	4.76
Number of points	141,732	150,320	114,021	137,424	144,107	184,773
Number of invalid points	7,253	6,743	10,059	2,464	10,072	3,699

percentage of correctly labelled voxels (bone vs non-bone when comparing the automatic segmentation to the ground truth segmentation) to total number of voxels.

As voxels in CT images generally contain much more non-bone than bone, it may be possible to have a high DSC score while misidentifying a significant number of the bony voxels. Because of this, an important metric in determining the accuracy of voxels is specificity. Specificity is defined as the number of true positives divided by (number of true positives plus the number of false positives). In this case, specificity is the number of voxels correctly identified as bone divided by the number of voxels correctly identified as bone summed with the number of voxels incorrectly identified as bone. The average specificity for these cases was 0.92.

The ASD results for the six segmentations are listed in Table 3.4. ASD is the measure of error when comparing the segmentation surface to the ground truth surface on a 3D point-wise basis. The total average surface distance for the segmentations was 2.17 mm, using 872,377 points. A discussion about these results is located in Section 4.2.1.

Certain artefacts often exist within the CT data (i.e. the CT table) which are later ignored by the main detection algorithm, as these artefacts are too far away to match well in the registration process. However, these artefacts can negatively affect segmentation results, although they do not usually affect final detection accuracy. Because of this, outlier points more than two standard deviations away were ignored when calculating ASD as they are also ignored within the registration process. These artefacts were not removed when calculating the DSC metric so the true accuracy may be slightly higher.

3.3.2 CBM considerations

As the 18 mm length can be adjusted as needed, an experiment was performed to see if any lesions were located at a depth greater than 10.8 mm away from all points in the segmentation. Lesions in this state would be too deep to be detected using this line length. This test was carried out on all diseased data sets and it was found that 3 lesions out of 1568 were too deep to detect.

In these three cases, at least one vertex normal from the bony surface intersected with a lesion at a depth greater than 10.8 mm. However, not a single ray did so within the 10.8 mm. This justifies the original use of 18 mm as the line length for CBM.

Although using a line length greater than 18 mm could ensure all lesions fall within the detection range, this will produce slower running times during CBM and is largely unnecessary as only a few lesions were found to be too deep in this data set.

It may also be possible, if the density of points on a bony surface mesh is too small in comparison to the size of the surface of the lesion, for all surface normal rays to miss intersecting with the lesion, even if the lesion is within 10.8 mm of the surface. A test found that this case was present with a single lesion. This justifies using the chosen resolution "high" in the Stradview software to create the segmentation mesh, as only one lesion was missed out of many. Using a polygonal mesh with a higher density of points could solve this problem. However, using a greater number of points will produce a slower run time and a greater memory cost, which, again, is unnecessary.

3.3.3 Validation of registration

Calculating registration error directly is a challenging problem, as it is difficult to create an accurate ground truth mesh that has been deformed from the original mesh into a completely correct shape. It is not clear how to define a registration that will move each specific preregistered point into its perfect location. It may be the case that multiple good solutions exist for a point, or that a single satisfactory solution does not even exist. Also, as the meshes in use contain hundreds of thousands of points, this problem is magnified by the sheer size of the point meshes.

It may be possible to place landmarks by hand at distinct point locations across both the preregistered mesh and the ground truth mesh, and then to compare the final landmark distances. However, this would only check the registration accuracy at these landmark points. There is no guarantee that the deformable registration method will prevent other points from moving away from the distinct locations during registration, even if the final registration of the landmark points is very accurate.

Because of these difficulties, the residual registration error was used to validate the registration. In this method, the atlas was allowed to deform and register to a segmentation using the same deformable registration technique used to project the symmetrical mapping onto the segmentation as described in Section 2.4. Whilst the residual error will not measure incorrect registration within the direction of the surface, it will still capture the major component of error where the surfaces are relatively complex and the registration is reasonably constrained. In this case we are dealing with the whole skeleton, which is undeniably a complex structure, and the registration is never allowed to deform too far in any direction (due to the smoothing process between each iteration).

In this experiment, the registration method was tested to see if each point in the atlas could successfully deform so that all points in the segmentation mesh closely match (using Euclidean distance) with a point in the atlas mesh. This experiment does not directly check whether the registration is correct for each point, but instead finds the average surface distance between each point in the segmentation mesh and its respective closest point in the deformed atlas mesh. Although this is not measuring the direct amount of registration error, this is a valid test which produces important results.

As a segmentation can often extend far below (in a negative z direction) or above (in a positive z direction) the atlas, segmentation points that lie outside the minimum z and maximum z atlas bounds will not have accurate registration matches. As these points will be ignored later in the detection algorithm due to poor registration matching and will not affect the overall detection results, they are ignored in this calculation by limiting the segmentation points to be within the maximum and minimum z bounds of the atlas. Figure 3.7 shows an example of a case where the segmentation extends below the atlas.

Other causes of poor registration are cases that contain either the CT imaging of a contrast agent administered to the patient or cases where the CT table is contained within the segmentation. Figure 3.7 shows an example of a case where the CT table is part of the segmentation. These issues, again, do not generally affect the detection algorithm's accuracy due to poor registration matching (these points are ignored), but only cause inaccurate and poor registration test results. Because of this, the three segmentations with contrast agents and the six segmentations containing CT tables were not used in this experiment.

Furthermore, another issue exists (in five cases) where the ribs are at significantly different longitudinal axis locations due to strong differences in the natural anatomy of the patient when compared to the anatomy of the atlas. This severely decreases the registration accuracy as the atlas ribs register to the segmentation vertebrae leaving the segmentation ribs without decent registration matches. This problem occurs as the ribs are not rooted to anything but are free to move a significant distance away from their location of origin if they do not have

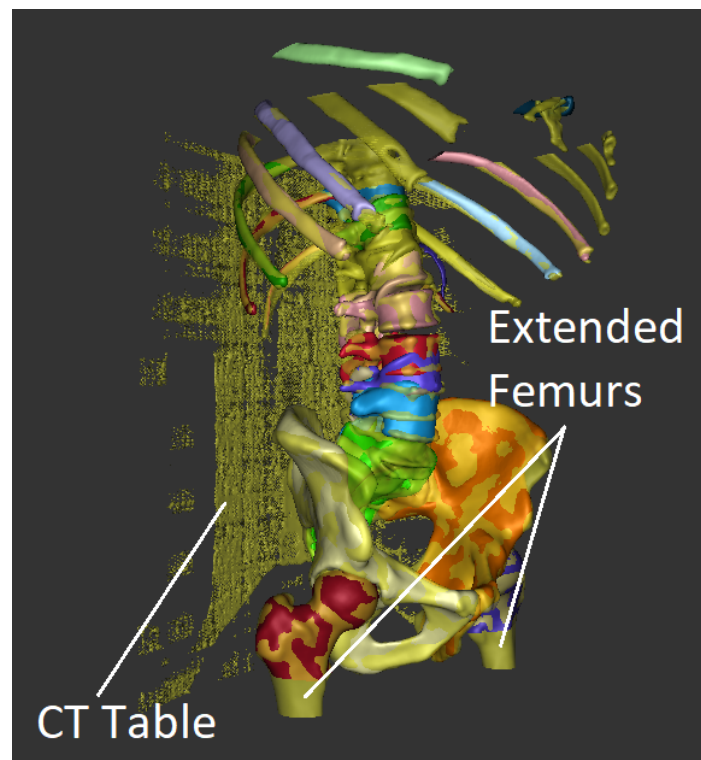
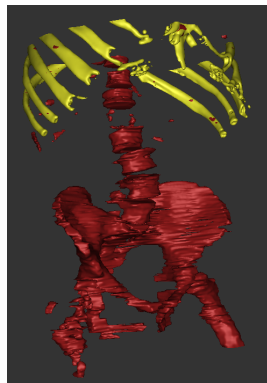
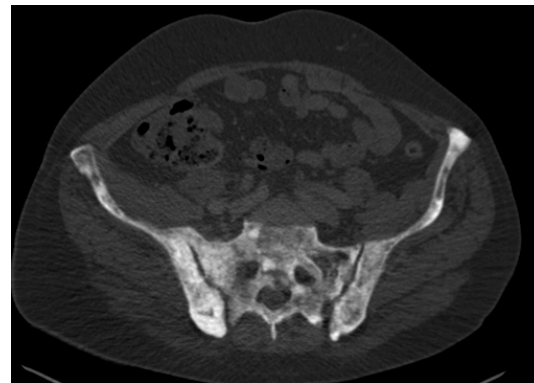


Fig. 3.7 An example of a segmentation (shown in yellow) that includes the CT table. The segmentation also extends far below the atlas femurs (shown in red and purple). Both of these issues can severely reduce the registration accuracy between the atlas (shown in multi-colour) and the segmentation. Note that in this case one of the ribs has also failed to register due to incomplete data in the segmentation.



(a) The lesions within this case are shown in red while the segmentation is shown in yellow. It can be seen that this case is extremely diseased.



(b) A CT slice image through the hips and pelvis for this case. The HU values for this case vary greatly from that of a normal CT scan.

Fig. 3.8 The images from a case of an extremely diseased patient. The segmentation algorithm achieved poor results on this case as the spine, pelvis, and femurs contain very different HU values from that of a healthy CT scan. As this case would be very easy for a radiologist to find the lesions, it is not a good candidate for this algorithm and is not used in the experiments. The CT data is from the Addenbrooke's Hospital CT Data Collection.

a good initial registration match. As there are relatively few lesions that occur in the ribs (Figure 3.4) and the thickness of the ribs are the most often too thin for CBM method to work (as described in Section 3.3.2), these five cases were removed from this experiment, so as not to confuse the registration error results for the other cases. However, these cases are used in the assessment algorithm and have a poor accuracy.

It was also found that one case was so completely diseased that the segmentation algorithm produced extremely poor results as the HU values were not representative of a normal CT scan. This case is shown in Figure 3.8. As this case was so obviously diseased, it would most likely not be necessary to run a detection algorithm to find the lesions within it. It was hence removed from all future experiments below, including in this registration experiment.

The registration results for the remaining 88 cases are listed in Table 3.5, summarised by bony region – that is, by the bony region within the atlas that a target point was closest. The average distance error between points was 4.84 mm with a standard deviation of 10.30 mm. A significant part of the registration error occurs in the ribs (for the same reasons described above). If ribs are discarded in this registration experiment, the average distance error between points was 3.38 mm. A discussion about these results is located in Section 4.2.2.

Table 3.5 The average distance between each point in the segmentation and its closest respective point in the atlas, for 88 cases. The pelvic bone is made of the ilium, acetabulum, ischium and pubis bones.

Region	Average distance (mm)	Standard deviation
femur left	2.79	3.47
femur right	2.82	3.61
sacrum	2.70	2.48
pelvic bone left	1.78	2.41
pelvic bone right	1.87	2.65
6th rib left	9.91	8.98
6th rib right	7.60	7.28
7th rib left	4.09	8.51
7th rib right	3.41	5.49
8th rib left	3.34	5.93
8th rib right	4.46	8.14
9th rib left	3.79	7.12
9th rib right	3.31	5.87
10th rib left	3.83	6.65
10th rib right	3.75	6.21
11th rib left	4.42	7.20
11th rib right	4.27	6.91
12th rib left	7.69	10.25
12th rib right	6.69	9.00
Th 10	8.14	9.13
Th 11	2.67	2.89
Th 12	2.59	2.77
L 1	2.23	3.14
L 2	1.91	2.33
L 3	1.94	2.28
L 4	2.15	2.53
L 5	2.38	3.01

3.4 Experiments on symmetrical disease assessment

3.4.1 Defining metrics

In the following experiments, a number of metrics were used to determine the quality of the assessment when comparing the results to the ground truth markings. The metrics listed below were selected as they provided the best overlap with those used by existing studies.

Existing studies provide a 2D point accuracy, so the point accuracy was included as a metric in this thesis (although these are 3D disease shadow points and significantly different from the 2D images used by the other studies). These studies also include a sensitivity metric, so this, too, was included. Additionally, a group accuracy was included in the thesis as a way of measuring the cluster size parameter's ability to remove erroneous regions from the results.

Point accuracy was determined by dividing the total number of correctly labelled points by the total number of points. Diseased point accuracy was found by dividing the number of points correctly labelled as disease by the total number of points labelled as diseased.

Sensitivity was found by comparing the points labelled as disease by the assessment with the points in the disease shadows produced by the ground truth 3D lesions. If at least a single point in the disease shadow was labelled as diseased by the assessment, the 3D lesion was marked as detected. The sensitivity was set to be the number of detected lesions divided by the total number of lesions.

Likewise, group accuracy was found by considering the individual groups of points (connected by their triangle point relations) marked as diseased by the assessment method. A disease group was considered correctly identified if at least a single point within the group was actually diseased. The group accuracy was set to be the number of correctly identified diseased groups divided by the total number of diseased groups.

3.4.2 Overall results

Once each vertex in every segmentation was labelled as being either diseased or healthy by the ground truth markings (as described in Section 3.2.3) and by the symmetrical disease assessment (as described in Section 2.5), the overall point accuracy and lesion detection sensitivity was determined. The disease side selection method described in Section 2.5.3 was also used to select diseased regions. It was found that the symmetrical disease assessment method produced an average sensitivity of 0.871 per case with a point accuracy of 0.894.

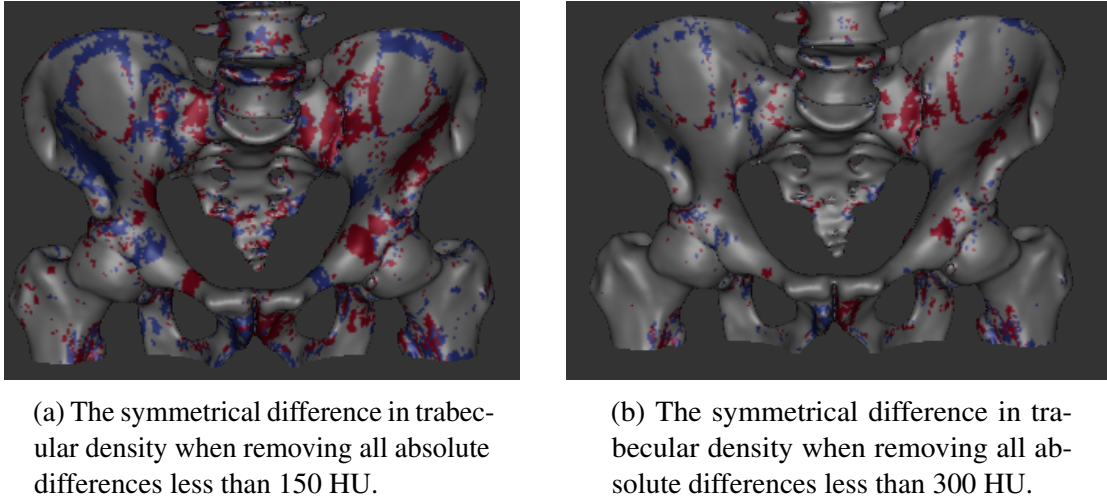


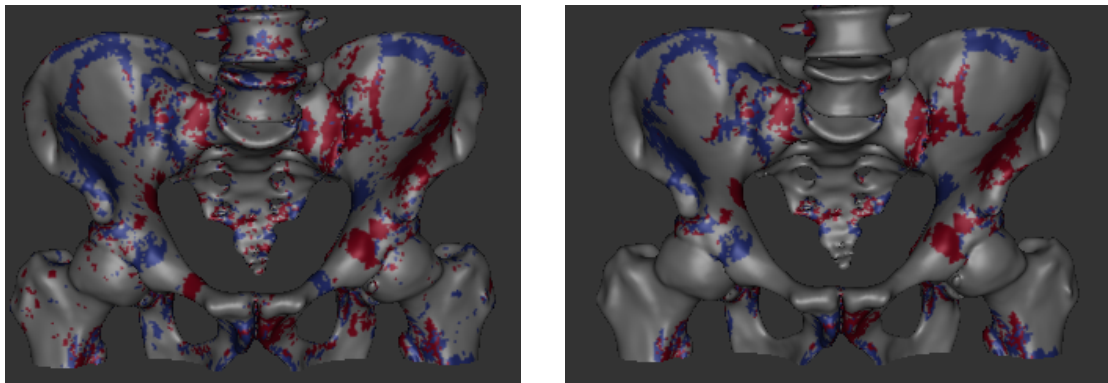
Fig. 3.9 These images show an example of how varying the threshold the HU difference cutoff can produce different results. The blue and red points are potentially diseased areas. The blue regions indicate positive differences greater than the cutoff value while red regions indicate negative differences greater than the cutoff value. Increasing the threshold HU cutoff will increase the amount of difference required in order for a vertex to be labelled as diseased.

3.4.3 Parameters

In order to achieve these results using the symmetrical disease assessment method, three main parameters were varied. The first of these was the threshold at which the absolute value of the symmetrical difference in radiodensity (HU) causes a set of symmetrical points to be labelled as diseased. Figure 3.9 shows the results when using two different thresholding values. A lower value will produce a more sensitive detection but will also incur more false positives, whilst a higher value will decrease sensitivity but also decrease the number of false positives.

The second parameter that can be varied, is a connected component size threshold filter. By treating the polygonal mesh as an undirected graph, where vertices are the nodes, and triangle relationships between these nodes define the paths, connected groups of diseased points can be grouped into clusters. This size parameter can be used to disregard all clusters where the number of points within that cluster is less than the selected cluster parameter. This is demonstrated in Figure 3.10. Often, the number of points in a diseased region is larger than the number of points in a group of points incorrectly labelled as disease – so this approach can help reduce error. However, there is always the risk of using too great a cluster size parameter as this can filter out small clusters of legitimate disease.

The third parameter that can be adjusted is the number of smoothing iterations used. This smoothing filter is described in Section 2.5.1. As small groups of speckled noise usually



(a) Significant groups of error exist when comparing the symmetrical difference in trabecular bone density.

(b) By removing clusters of points less than a defined threshold, much of the error can be disregarded.

Fig. 3.10 These images show an example of how using a cluster size of 300 can be used to remove speckled noise.

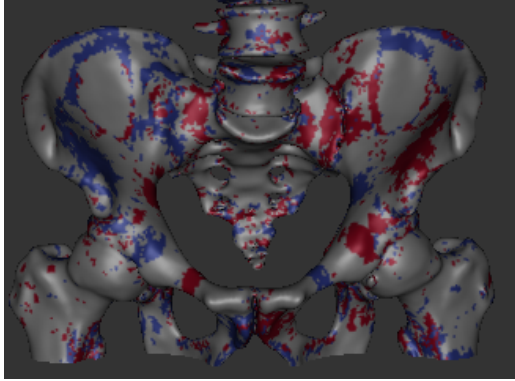
exist, this parameter can help reduce error as this technique removes the less precise CBM measurements by smoothing the trabecular bone density measurements throughout the mesh. This is shown in Figure 3.11. However, smoothing can also remove legitimate disease from the results, decreasing the sensitivity.

3.4.4 Determining parameters

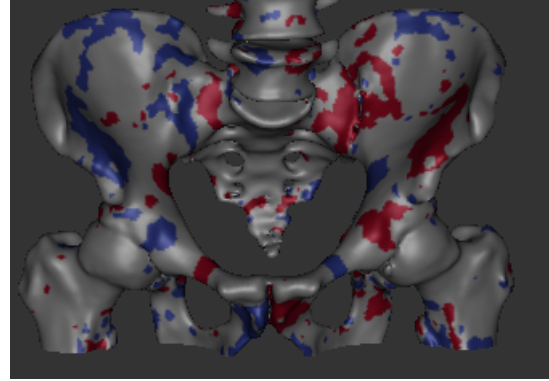
To determine optimal values for the three parameters, a brute force method was first used to evaluate a wide range of sensible combinations of parameters. This involved trying over 2000 unique parameter combinations on the 102 data sets. (One data set out of the original 103 cases was removed as it was completely diseased. This data set was described in Section 3.3.3)

For this experiment, the HU cutoff threshold was varied from 25 to 400 HU, the number of smoothing iterations was varied from 0 to 20 and the cluster size was varied from 1 to 20. Results were sorted by multiple evaluation metrics as described below.

Table 3.6 lists the best 10 results sorted by point accuracy. As shown in the table, the best overall point accuracy was found by using high values for all of the parameters (HU cutoff threshold, the number of smoothing iterations and the cluster size). As the majority of points in a data set are healthy points, using high parameter values decreased the number of points marked as disease. Because of this, the point accuracy increased but the sensitivity decreased.



(a) Significant speckled noise exists when comparing the symmetrical difference in trabecular bone density.



(b) By smoothing for a number of iterations, much of the speckled noise can be removed, improving accuracy.

Fig. 3.11 These images demonstrate how using three smoothing iterations can remove speckled noise.

Table 3.6 The 10 best results and parameter values when sorting by point accuracy using a brute force method.

HU threshold	Smoothing iterations	Cluster size	Point accuracy	Average sensitivity
400	20	5	0.9537	0.1529
400	20	6	0.9537	0.1487
400	20	15	0.9539	0.1219
400	20	7	0.9537	0.1450
400	20	4	0.9536	0.1539
400	20	8	0.9538	0.1405
400	20	9	0.9538	0.1358
400	20	20	0.9540	0.1139
400	20	3	0.9536	0.1566
400	20	10	0.9538	0.1313

Table 3.7 The 10 best results and parameter values when sorting by sensitivity using a brute force method.

HU threshold	Smoothing iterations	Cluster size	Point accuracy	Average sensitivity
25	0	1	0.747	0.978
25	0	2	0.755	0.977
25	0	3	0.760	0.976
25	1	1	0.747	0.975
50	0	1	0.791	0.974
25	0	4	0.764	0.973
25	1	2	0.756	0.972
25	0	5	0.767	0.971
75	0	1	0.823	0.970
25	0	6	0.770	0.970

Table 3.7 displays the best 10 results sorted by sensitivity. The greatest sensitivity was found using low values for the parameters. By decreasing the values for these three parameters, more points were marked as diseased, which improved the sensitivity but increased the number of false positives.

Maximising point accuracy produced lower sensitivities, while maximising sensitivity produced more false positives. In order to produce usable results, parameter values that are influenced by both point accuracy and sensitivities were needed to balance the sensitivity with the number of false positives.

In order to find this balance, the results were sorted by both point accuracy and sensitivity. To combine accuracy and sensitivity, values were either summed or multiplied together. Different weightings and normalisations were applied by multiplying each value by a selected weight from a range of weightings.

A more systematic method of normalisations was also tried by dividing each term by its respective squared variance or by applying a linear weighting scale (equation 3.4). This was based on a value's relation to the maxima or minima across the data sets. These two terms were added as in equation 3.5 and then sorted. In this equation, N_a is the normalised accuracy and N_s is the normalised sensitivity.

$$normalisation = \frac{value - minima}{maxima - minima} \quad (3.4)$$

$$combination = N_a + N_s \quad (3.5)$$

Table 3.8 shows the results of using this equation to combine terms. This same process was repeated by multiplying the terms instead of summing them and by maximising many

Table 3.8 The 10 best results and parameter values when sorting by point accuracy and sensitivity using a brute force method. The point accuracy and sensitivity were first normalised and then combined into a single term. These results were then summed and sorted.

HU threshold	Smoothing iterations	Cluster size	Point accuracy	Average sensitivity
125	0	1	0.865	0.962
175	0	1	0.891	0.925
200	0	1	0.900	0.911
125	1	1	0.874	0.947
225	0	1	0.907	0.901
125	0	2	0.874	0.946
150	1	1	0.889	0.925
275	0	1	0.918	0.882
250	0	1	0.913	0.889
175	1	1	0.900	0.904

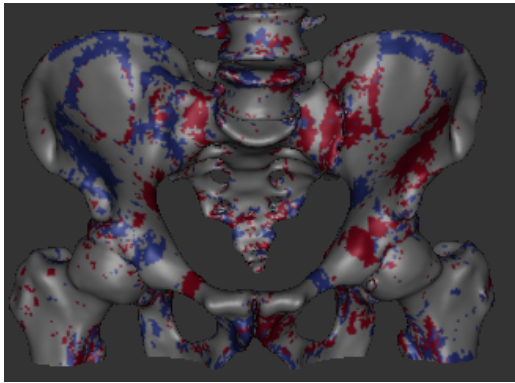
other combinations of metrics including using diseased point accuracy, specificity and group accuracy. Combining more than two terms was also tried but it was found that all of these variations and combinations produced similar results. The results from these experiments did not converge on a set of optimal parameters, but showed that there is a trade off between sensitivity and the number of false positives.

3.4.5 Parameter considerations

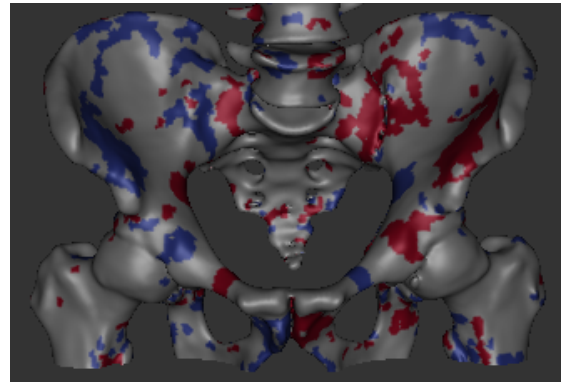
Although using a brute force method to maximise the different metrics was useful to gain a general understanding of how the different parameter values affected the results, it was found that there are subtle lower bounds on the parameters that need to be in place in order to produce usable and valid results. These bounds are difficult to find by only examining the numerical results and can be recognised more easily through visual results.

When using low parameter values, a speckled error noise pattern (of disease points) exists that semi-evenly covers the bony surface and greatly increases the sensitivity rate. The point accuracy does not drastically decrease due to the relatively low density of the speckled pattern. This speckled noise can be seen in Figure 3.12a. Removing speckled error is essential to producing meaningful results as blindly marking evenly spaced points on the bony surface as diseased will produce a high sensitivity rate with a relatively low number of false positives. If this speckled error is not removed, these results will be useless in practical detection of lesions.

One way to reduce this speckled noise error is to set the number of smoothing iterations to at least two and to set the minimum cluster size to be at least seven. These numbers were



(a) A speckled pattern of error exists when using low smoothing and low cluster sizes. This speckle must be removed for meaningful results.



(b) By applying a minimum cluster size of seven and two smoothing iterations, most of the speckle noise can be removed.

Fig. 3.12 These images show an example of how using a number of smoothing iterations and a minimum cluster size can reduce speckled noise.

found through a visual inspection of the results. Figure 3.12 shows the results of using 2 smoothing iterations and a minimum cluster size of 7.

From Section 3.4.4, a valid cost function that could be optimised does not appear to exist. However, there is an appropriate range of values that could be used as input parameters to the algorithm that would achieve satisfactory results. This range generally includes a minimum of two smoothing iterations and a minimum cluster size of seven—though a range of values greater than this would be sufficient.

3.4.6 Overfitting

By training and testing the optimised parameters on a single data set, it is possible that the parameters will over-fit (overfitting is when an optimised model performs well on the training data, but does not perform accurately on a different set of data). In order to help prevent overfitting, the data can be split into a training and a testing set. The training data set is used to optimise the parameters and hyperparameters while the testing data set is used to validate the accuracy of these optimised parameters in an unbiased way [73].

The most common method of splitting the data is to place 80% of the data into the training set and 20% of the data into the testing test. The training set is then split again into data that is used to train the model and data that is used to validate the model. General splitting methods for this include cross-validation, the bootstrap method, and selecting the most representative samples for the split [158]. The simplest form of data splitting is cross-validation (CV), which involves dividing the data set into the two groups at random - usually with 80% of the

data used for training and 20% used for validation. There are many variations on this which include the leave-one out CV, Monte-Carlo CV and K-fold CV.

In leave one out CV, a single data set is taken out of the training set and used for validation on each iteration. This is repeated for each item in the data set until every data set has been left out one time. In Monte-Carlo CV, the data is selected at random and placed into either the training set or into the validation set. This is repeated for a number of iterations. Usually the percentages differ between the training set and validation on each iteration. The same data sets can be placed into the training or validation sets on subsequent iterations. In K-fold CV, the data is split randomly into K groups (folds). On each iteration, one fold is used as the validation group and the rest of the folds are used as the training group. This is repeated K times, until each fold has been used as a validation group. In the bootstrap method, samples are iteratively selected from a data set with replacement into the training or validation sets [39]. Other methods include selecting representative samples and placing them into the training set or validation set.

Although all of these methods could be used successfully to prevent overfitting, K-fold cross validation was used in this project for a number of reasons. Using K-fold CV is the gold standard of data splitting. Using K-folds is much more robust than using a single fold for validation. The other methods are more complicated and are more computationally expensive while not providing many extra benefits. They would most likely produce similar results. The most common split is an 80% to 20% (training to testing) split. However, many other similar ratios are commonly used (i.e. 70% to 30% or a ratio of 1/3 to 2/3 [73]). Any ratio similar to these percentages will work successfully.

In my project, all of the data was originally placed into the training set and none was left for the testing set. It is too late to create a testing set now without acquiring more data (as all of the data has been used as part of the training). The lack of a testing set is a limitation, as a testing set would have provided an unbiased evaluation of the optimised parameters. However, a 5-fold validation technique was used and this does provide a robust measure of the accuracy (although it may be slightly biased). All of the experiments described below this section use the 5-fold CV technique for training and validation.

3.4.7 Parameters based on volume size

As shown in Table 3.1, the majority of the lesions in this data set have a volume less than 3.00 cm³. However, lesions that have a volume in the range between 3.00 cm³ and 6.00 cm³, produced an average disease shadow size of 597 points. The minimum cluster size parameter can be successfully set to a much higher value when detecting larger lesions than when detecting small lesions.

Related to this, it was also found that the disease shadow produced by the symmetrical assessment method has significantly fewer points than that of a disease shadow produced by the ground truth lesions. This may be because the ground truth size of the disease shadow is the sum total of all of the disease shadows produced by a lesion while the symmetrical assessment method assesses each disease shadow individually. As the cluster size needs to be reflective of this difference, it must be adjusted accordingly.

To determine parameters based on the lesion volume size, a number of experiments were run. The sensitivity and accuracy results of these experiments were binned into different lesion volume groups. In Figures 3.13a, 3.13b, 3.13c, 3.13d and 3.13e, the HU threshold cutoff was set to 150, the number of smoothing iterations was set to 2 and the cluster size was allowed to vary. Different folds from the training data were used to produce each result. These figures show that as the minimum cluster size increases, the sensitivity decreases. Figure 3.14 shows that as the cluster size increases, the overall point accuracy increases. As the cluster size is decreased, more lesions will be detected at the cost of a reduced accuracy.

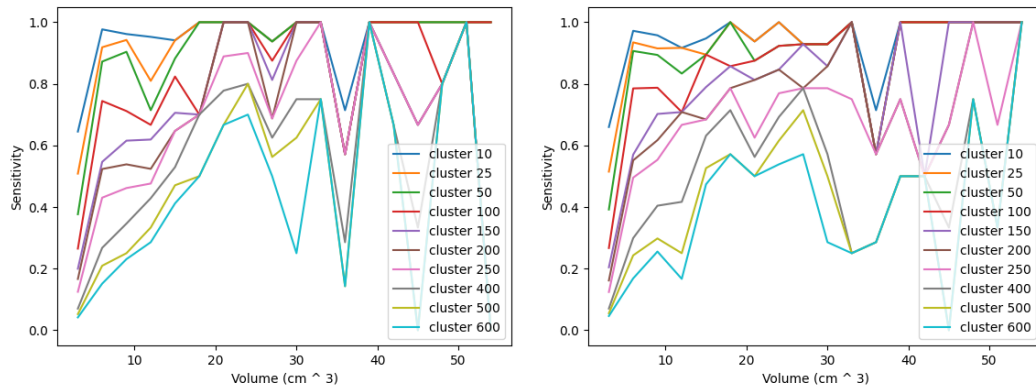
In the next experiment, the HU threshold cutoff was allowed to vary, the number of smoothing iterations was set to 2 and the cluster size was set to 7. Figures 3.15a, 3.15b, 3.15c, 3.15d, and 3.15e show the results of using these parameters while Figure 3.16 shows the accuracy using these parameters. Higher HU thresholds can be used to produce better accuracies at the cost of lower sensitivities and vice versa. This result is similar to the clustering size result. Lesions that are detected using either a high cluster size or a high threshold will be a more accurate detection. However, it is not necessarily the case that a lesion has increased in trabecular density as it becomes more diseased or becomes larger.

The experiment was also repeated for the smoothing iterations. The HU threshold cutoff was set to 150, the cluster size was set to 7 and the number of smoothing iterations was allowed to vary. It was found that using 2 to 5 smoothing iterations improved results, but varying the cluster size was more successful than varying the smoothing iterations.

In Figure 3.17 and Figure 3.18, lesions have been binned by a volume cutoff (i.e., all volumes greater than the cutoff are placed in the same bin) for the varying cluster and threshold data. A single volume can be placed into multiple bins.

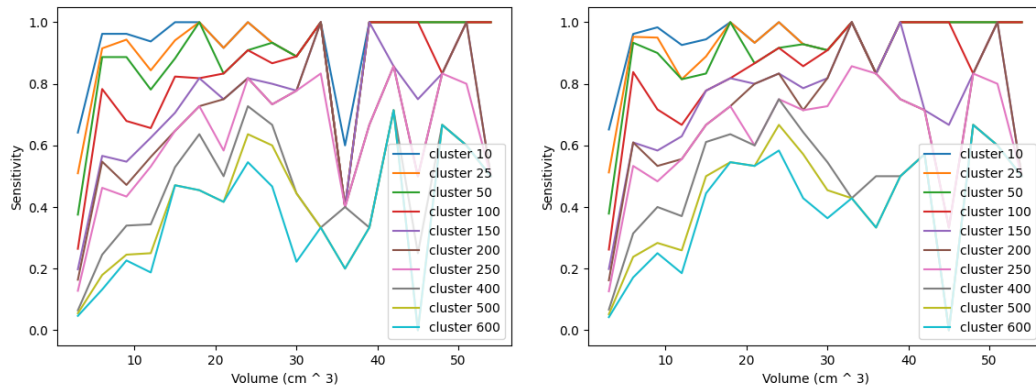
3.4.8 Overall optimised results

As there is no cost function to optimise, the graphs above (Figures 3.13a, 3.13b, 3.13c, 3.13d and 3.13e, 3.15a, 3.15b, 3.15c, 3.15d, 3.15e, 3.17 and Figure 3.18) were used to select appropriate parameters for these optimal results. These parameters were generated by their respective CV training data sets. The best results use the same set of parameters – a fixed cluster size of 7 with 2 smoothing iterations and a threshold cutoff from 100 HU to 150



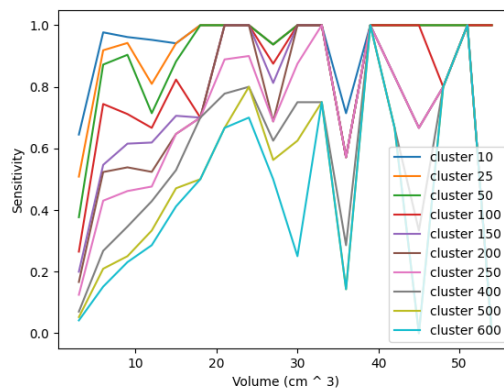
(a) Training data excluding fold one data was used to produce this graph.

(b) Training data excluding fold two data was used to produce this graph.



(c) Training data excluding fold three data was used to produce this graph.

(d) Training data excluding fold four data was used to produce this graph.



(e) Training data excluding fold five data was used to produce this graph.

Fig. 3.13 These graphs show the relationship between sensitivity, volume and cluster size. In this experiment, the threshold cutoff size was set to 150, the number of smoothing iterations was set to 2 and the cluster size was allowed to vary. The lesions were binned into 20 groups by ranges of volume. As the cluster size increases, the sensitivity decreases.

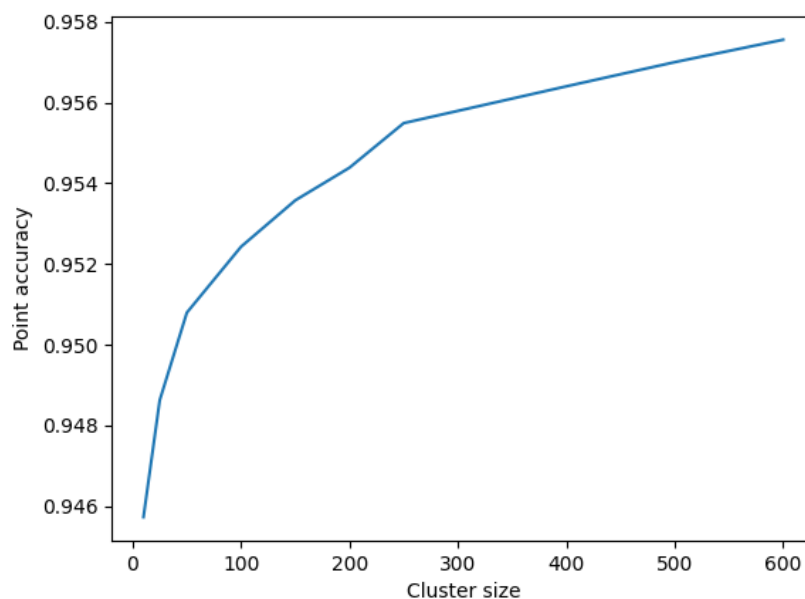


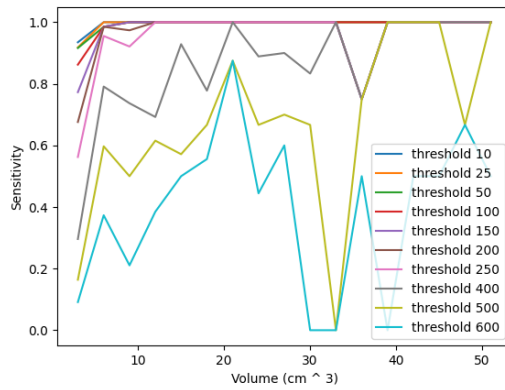
Fig. 3.14 A graph showing the cluster size vs point accuracy using the training data that does not include data from fold one. As the cluster size is increased, so does the overall point accuracy.

HU. This fulfils the minimum thresholds needed to remove the speckled noise, while also maximising the sensitivity. These values can be increased to improve accuracy at a cost of a decreased sensitivity. A discussion about the parameter choice is located in Section 4.2.3. Tables 3.9, 3.10 and 3.11 show the results of using these optimal parameters in detecting mixed lesions on the validation data sets.

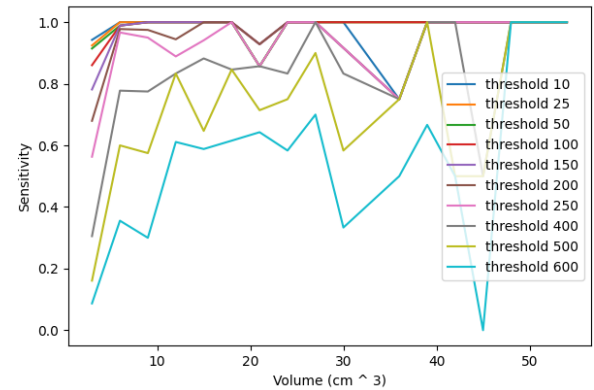
When using a HU threshold cutoff of 100 HU, the resulting average point accuracy and average sensitivity was 0.874 and 0.867 per case respectively. When using a HU threshold cutoff of 125 HU, the resulting average point accuracy and average sensitivity was 0.894 and 0.871 per case respectively. When using a HU threshold cutoff of 150 HU, the resulting average point accuracy and average sensitivity was 0.908 and 0.808 per case respectively.

A discussion about these results is located in Section 4.2.4. Examples of typical good results are shown in Figure 3.20, Figure 3.21 and Figure 3.22. Figure 3.19 shows an example of a typical average result while Figure 3.23 shows an example of a typical poor result.

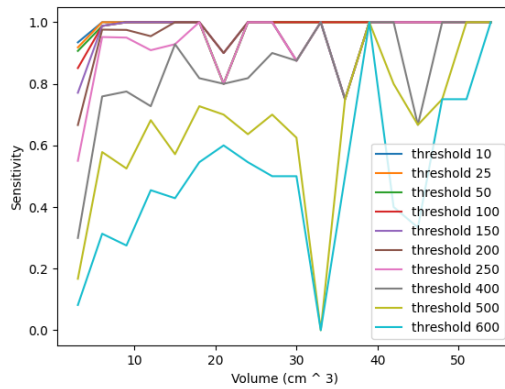
Using the results from section 3.2.4, for a cluster size of 7, the total surface area was calculated to be 28.63 mm². This brings the rough surface area of a cluster of 7 points to 20.45 mm² as two points can be subtracted from the calculation as it contains a small number of points. This is about a 4.5 mm by 4.5 mm region of bone.



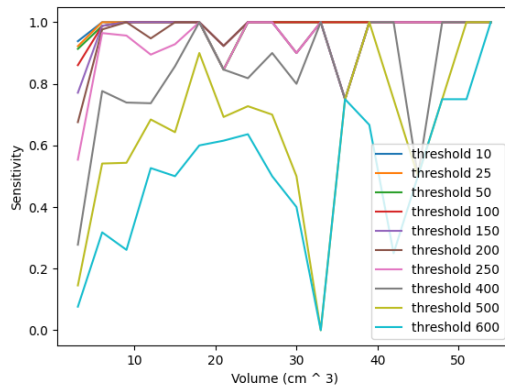
(a) Training data excluding fold one data was used to produce this graph.



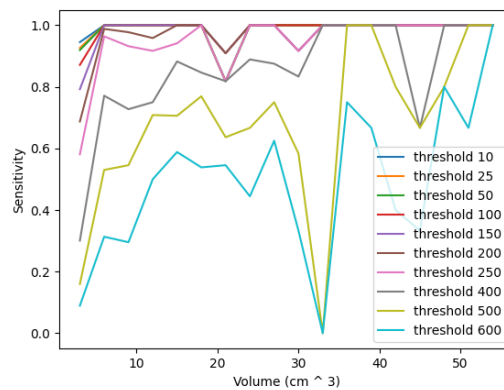
(b) Training data excluding fold two data was used to produce this graph.



(c) Training data excluding fold three data was used to produce this graph.



(d) Training data excluding fold four data was used to produce this graph.



(e) Training data excluding fold five data was used to produce this graph.

Fig. 3.15 These plots show the relationship between sensitivity, volume and threshold. In this experiment, the cluster size was set to 7 and the smoothing to 2 iterations and the HU threshold cutoff was allowed to vary. The lesions have been binned by range of volume. As the threshold increases, the sensitivity decreases.

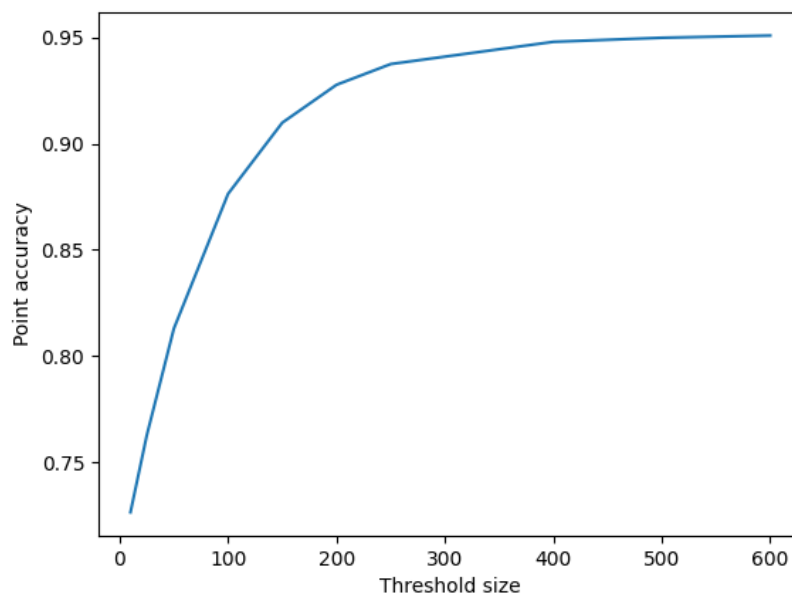


Fig. 3.16 A graph showing the threshold size vs point accuracy using training data that does not include fold one data. As the threshold size increases, so does the overall point accuracy.

Table 3.9 The average results per case from using the optimal parameters when detecting data sets containing both lytic and sclerotic lesions using the validation data sets. 2 smoothing iterations, a cluster size of 7 and a threshold cutoff of 100 were used.

Validation fold	Point accuracy HU	Average sensitivity HU
One	0.863	0.904
Two	0.870	0.899
Three	0.876	0.836
Four	0.875	0.886
Five	0.886	0.810
Average	0.874	0.867

Table 3.10 The average results per case from using the optimal parameters when detecting data sets containing both lytic and sclerotic lesions using the validation data sets. 2 smoothing iterations, a cluster size of 7 and a threshold cutoff of 125 were used.

Validation fold	Point accuracy	Average sensitivity
One	0.881	0.878
Two	0.890	0.899
Three	0.896	0.810
Four	0.893	0.860
Five	0.909	0.748
Average	0.894	0.871

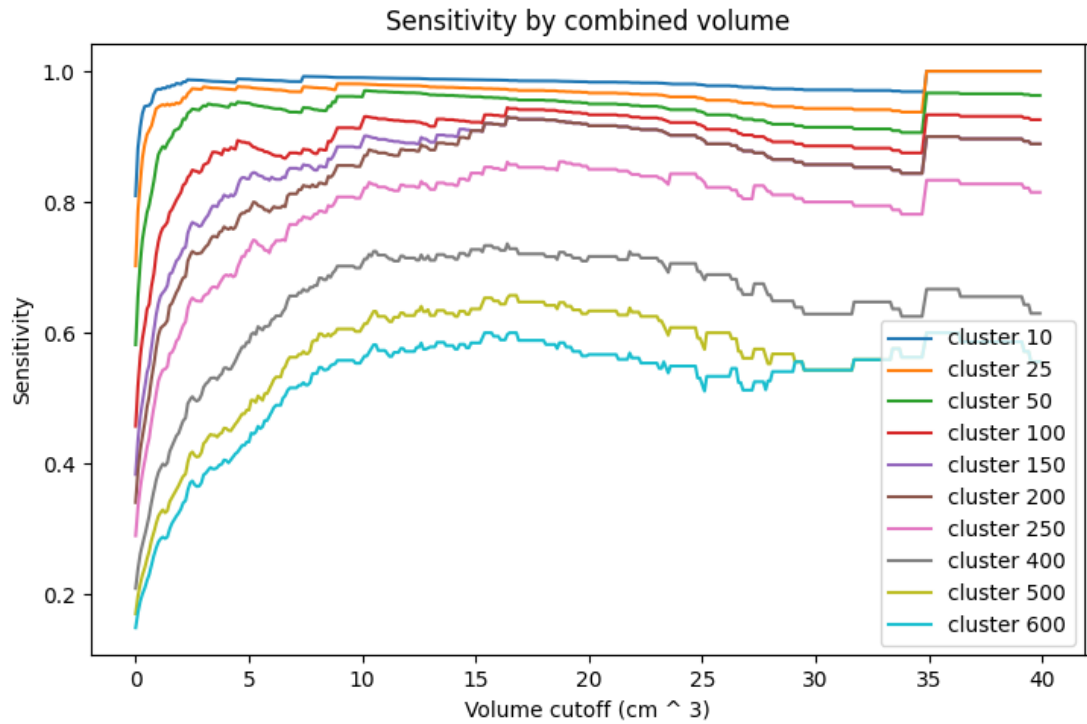


Fig. 3.17 A plot showing the relationship between sensitivity, volume cutoff and cluster size. Lesions have been binned by their volume cutoff. Each volume is placed into every bin where its volume is greater than the threshold cutoff. For this experiment, the HU threshold was set to 150 and the number of smoothing iterations was set to 2. This graph is intended to show how each cluster size affects the sensitivity of detecting lesions that are greater than a certain volume. It does not show how a cluster size affects the sensitivity of detecting a lesion of a certain volume. Training data that does not include fold data one was used to produce this graph.

Table 3.11 The average results per case from using the optimal parameters when detecting data sets containing both lytic and sclerotic lesions using the validation data sets. 2 smoothing iterations, a cluster size of 7 and a threshold cutoff of 150 were used.

Validation fold	Point accuracy	Average sensitivity
One	0.894	0.863
Two	0.905	0.883
Three	0.911	0.723
Four	0.907	0.855
Five	0.924	0.718
Average	0.908	0.808

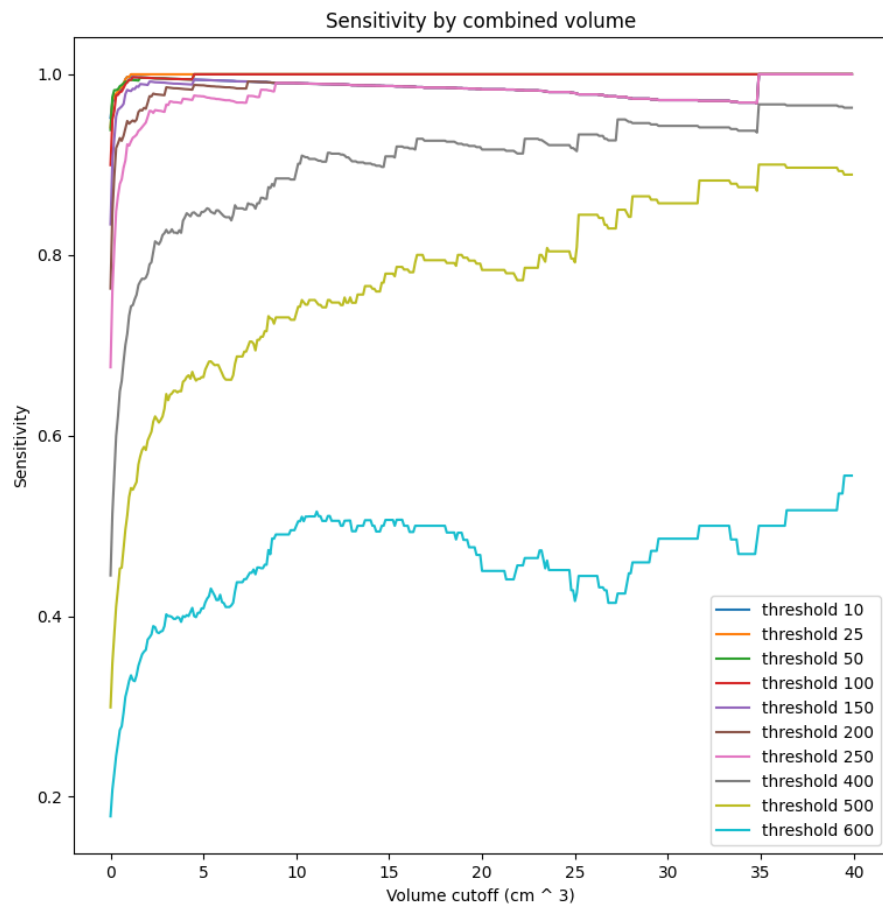
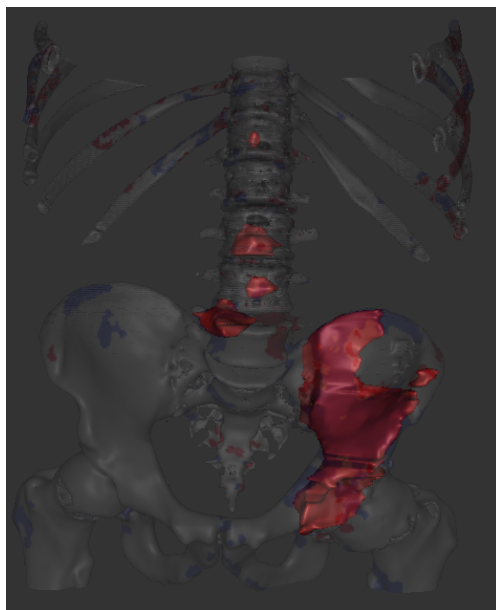
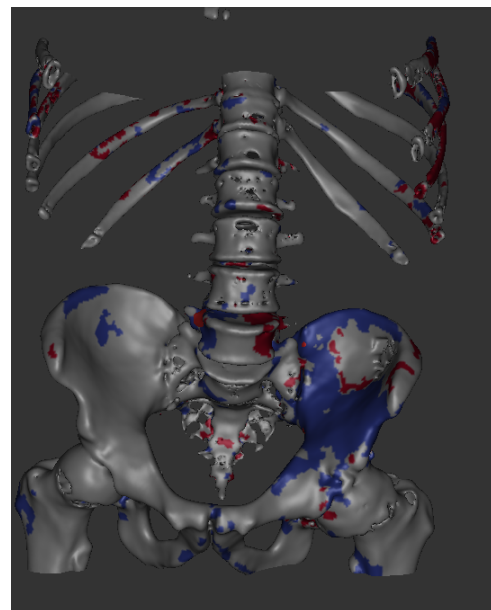


Fig. 3.18 A plot showing the relationship between sensitivity, volume cutoff and threshold. Lesions have been binned by their volume cutoff. Each volume is placed into every bin where its volume is greater than the threshold cutoff. The cluster size for this experiment was set to 7 and the number of smoothing iterations was set to 2. This graph is intended to show how each threshold affects the sensitivity of detecting lesions that are greater than a certain volume. Training data excluding fold one was used to produce this graph.



(a) Using an alpha blend to see the ground truth 3D volumes (shown in red) which are located within the bony surface.

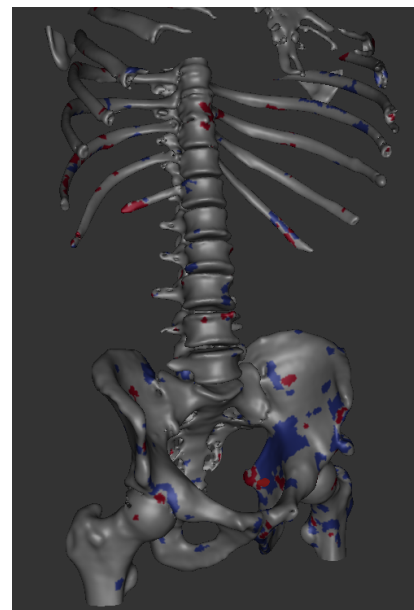


(b) Possible sclerotic lesions are shown in blue while possible lytic lesions are shown in red – found using the assessment method.

Fig. 3.19 A representative example of an average result from the fold five validation set. The assessment method found 5 out of 6 lesions correctly. It achieved a sensitivity of 0.833 with a point accuracy of 0.887.



(a) Using an alpha blend to see the ground truth 3D volumes (shown in red) which are located within the bony surface.

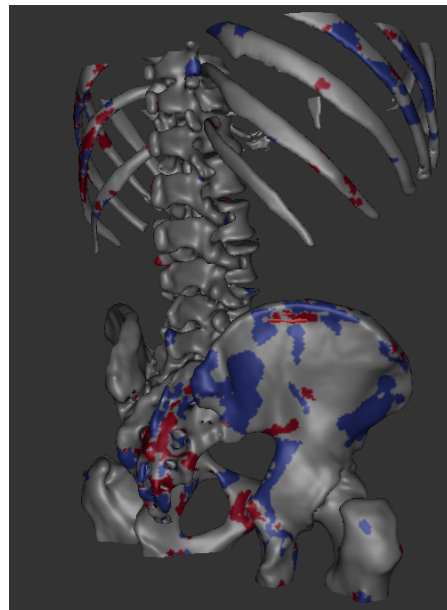


(b) Possible sclerotic lesions are shown in blue while possible lytic lesions are shown in red – found using the assessment method.

Fig. 3.20 A representative example of a good result from the fold five validation set. The assessment method found 23 out of 24 lesions correctly. It achieved a sensitivity of 0.958 with a point accuracy of 0.861.

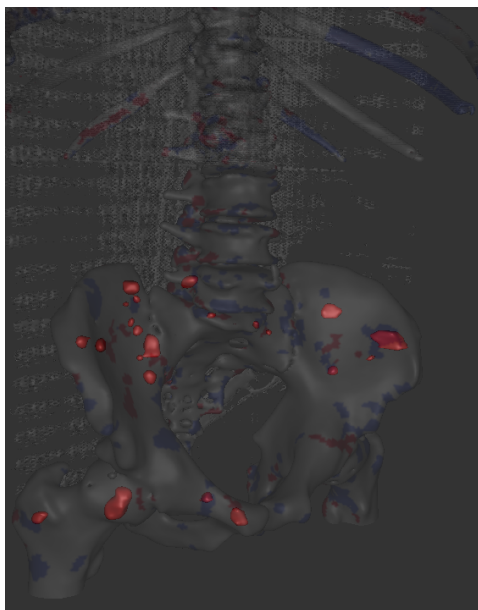


(a) Using an alpha blend to see the ground truth 3D volumes (shown in red) which are located within the bony surface.

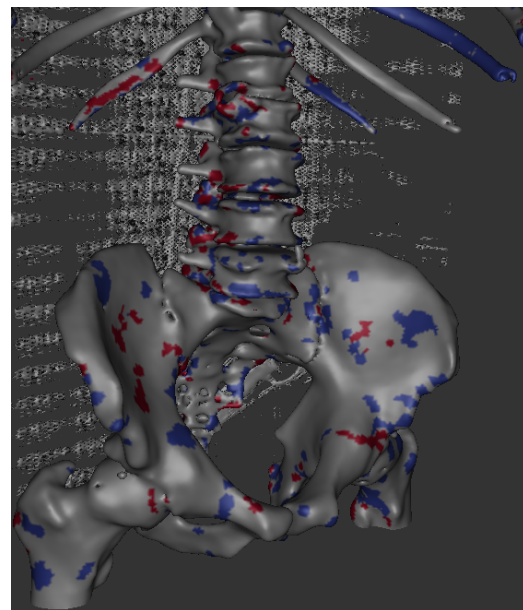


(b) Possible sclerotic lesions are shown in blue while possible lytic lesions are shown in red – found using the assessment method.

Fig. 3.21 A representative example of a good result from the fold five validation set. The assessment method found 23 out of 24 lesions correctly. It achieved a sensitivity of 0.958 with a point accuracy of 0.817.

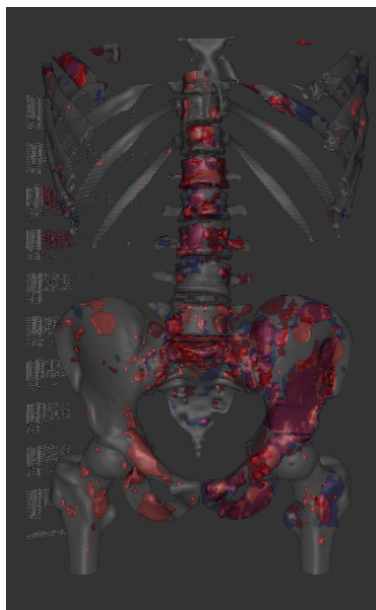


(a) Using an alpha blend to see the ground truth 3D volumes (shown in red) which are located within the bony surface.

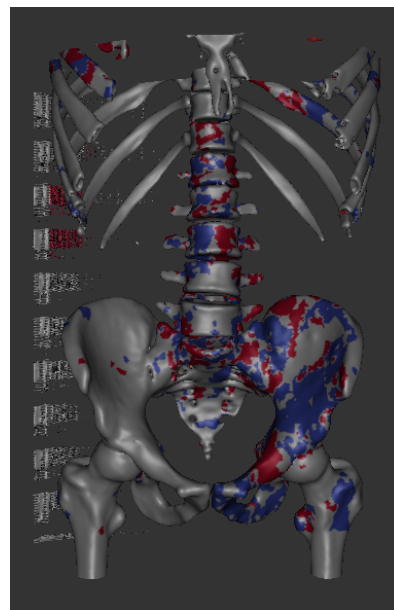


(b) Possible sclerotic lesions are shown in blue while possible lytic lesions are shown in red, found using the assessment method.

Fig. 3.22 A representative example of a good result from the fold five validation set. The assessment method found 19 out of 21 lesions correctly. It achieved a sensitivity of 0.905 with a point accuracy of 0.929.



(a) Using an alpha blend to see the ground truth 3D volumes (shown in red) which are located within the bony surface.



(b) Possible sclerotic lesions are shown in blue while possible lytic lesions are shown in red, found using the assessment method.

Fig. 3.23 A representative example of a poor result from the fold five validation set. This case achieved a sensitivity of 0.621 with a point accuracy of 0.823. The assessment method found 120 out of 193 lesions correctly.

As a representative example of validation, the results using data from the fold five data validation set was selected. At a point level, 48.5% of the ground truth diseased points were marked correctly (true positive rate) and 89.7% of the total number of healthy points were marked correctly as healthy (true negative rate). The false positive rate was 3.00%, which is (the number of false positives) / (the number of false positives + the number of true negatives). A total of 2,474,811 points were labelled correctly. Of this total, 70,084 were diseased points and 2,404,727 were healthy points. 74,460 diseased points were labelled incorrectly while 275,429 healthy points were labelled incorrectly. When limiting the results to only the 35 cases of healthy data, 91.91% of the points were marked correctly. The average false positive rate was 2.70% across all of the data sets.

At a diseased region level (i.e. a cluster of points marked as disease), a 30.6% true positive rate was achieved (the number of diseased regions with at least a single point that is actually diseased) / (total number of diseased regions). A false negative rate of 69.4% was achieved, which is the (number of regions incorrectly marked as disease) / (number of diseased regions).

In Figure 3.24, the number of missed lesions has been grouped by region. The lesions that were not detected had an average volume of 0.239 cm³ and an average disease shadow size of 34.6 points. Table 3.12 shows sensitivity by lesion volume while Table 3.13 shows the sensitivity by lesion volume where lesions have a volume less than 3.0 cm³. Table 3.14 shows the sensitivity by diseased shadow size and Table 3.15 shows the sensitivity by diseased shadow size where the size is less than 300 points.

3.4.9 Sclerotic and lytic labelling

As Dr. Barrett was unable to label individual lesions as being either sclerotic or lytic due to time constraints (and the Coronavirus pandemic), I developed a technique to identify these lesions correspondingly. In this technique, the trabecular bone density for each point in the ground truth disease shadow (that was produced by a ground truth lesion) was compared to the trabecular bone density of its symmetrically mapped point. The trabecular bone density was summed: the lesion was labelled sclerotic if its sum was greater than the sum of the symmetrically mapped points; if the sum was less than the sum of the symmetrically mapped points, the lesion was labelled lytic. The process is similar to the disease side selection technique described in Section 2.5.3. However, there are significant differences between these two methods as the disease side selection only compares border points on a region while the lesion labelling method compares the entire disease shadow.

This process labelled 273 lesions as sclerotic and 137 lesions as lytic. The assessment method detected 241 out of 273 sclerotic lesions correctly (0.883) and 72 out of 137 lytic

Table 3.12 The sensitivity of detecting all lesions by volume (cm³) using validation data from fold five.

Volume range (cm ³)	Number of detected lesions	Number of lesions	Sensitivity
$0 \leq \text{volume} < 3.0$	233	327	0.713
$3.0 \leq \text{volume} < 6.0$	28	29	0.966
$6.0 \leq \text{volume} < 9.0$	17	17	1.00
$9.0 \leq \text{volume} < 12.0$	2	2	1.00
$12.0 \leq \text{volume} < 15.0$	5	5	1.00
$15.0 \leq \text{volume} < 18.0$	3	3	1.00
$18.0 \leq \text{volume} < 21.0$	4	4	1.00
$21.0 \leq \text{volume} < 24.0$	5	5	1.00
$24.0 \leq \text{volume} < 27.0$	5	6	0.833
$27.0 \leq \text{volume} < 30.0$	1	1	1.00
$30.0 \leq \text{volume} < 33.0$	0	0	NA
$33.0 \leq \text{volume} < 36.0$	1	1	1.00
$36.0 \leq \text{volume} < 39.0$	0	0	NA
$39.0 \leq \text{volume} < 42.0$	0	0	NA
$42.0 \leq \text{volume} < 45.0$	1	1	1.00
$45.0 \leq \text{volume} < 48.0$	1	1	1.00
$48.0 \leq \text{volume} < 51.0$	1	1	1.00
$51.0 \leq \text{volume} < 54.0$	0	0	NA
$54.0 \leq \text{volume} < 57.0$	0	0	NA
$57.0 \leq \text{volume} < 60.0$	0	0	NA
$60.0 \leq \text{volume}$	6	7	0.857

Table 3.13 The sensitivity of detecting lesions where the lesion volume was less than 3 cm³ using validation data from fold five.

Volume range (cm ³)	Number of detected lesions	Number of lesions	Sensitivity
0 ≤ volume < 0.1	37	94	0.394
0.1 ≤ volume < 0.2	38	54	0.704
0.2 ≤ volume < 0.3	30	36	0.833
0.3 ≤ volume < 0.4	18	26	0.692
0.4 ≤ volume < 0.5	12	13	0.923
0.5 ≤ volume < 0.6	14	14	1.00
0.6 ≤ volume < 0.7	11	13	0.846
0.7 ≤ volume < 0.8	11	11	1.00
0.8 ≤ volume < 0.9	10	11	0.909
0.9 ≤ volume < 1.0	4	5	0.800
1.0 ≤ volume < 1.1	2	2	1.00
1.1 ≤ volume < 1.2	5	5	1.00
1.2 ≤ volume < 1.3	2	2	1.00
1.3 ≤ volume < 1.4	3	3	1.00
1.4 ≤ volume < 1.5	5	6	0.833
1.5 ≤ volume < 1.6	2	2	1.00
1.6 ≤ volume < 1.7	4	4	1.00
1.7 ≤ volume < 1.8	2	2	1.00
1.8 ≤ volume < 1.9	4	4	1.00
1.9 ≤ volume < 2.0	2	2	1.00
2.0 ≤ volume < 2.1	2	2	1.00
2.1 ≤ volume < 2.2	2	3	0.667
2.2 ≤ volume < 2.3	5	5	1.00
2.3 ≤ volume < 2.4	2	2	1.00
2.4 ≤ volume < 2.5	1	1	1.00
2.5 ≤ volume < 2.6	1	1	1.00
2.6 ≤ volume < 2.7	0	0	NA
2.7 ≤ volume < 2.8	1	1	1.00
2.8 ≤ volume < 2.9	1	1	1.00
2.9 ≤ volume < 3.0	2	2	1.00

Table 3.14 The sensitivity of detecting lesions by disease shadow size using validation data from fold five.

Disease shadow size range	Number of detected lesions	Number of lesions	Sensitivity
$0 \leq \text{shadow size} < 300$	222	319	0.696
$300 \leq \text{shadow size} < 600$	28	28	1.00
$600 \leq \text{shadow size} < 900$	26	26	1.00
$900 \leq \text{shadow size} < 1200$	5	5	1.00
$1200 \leq \text{shadow size} < 1500$	7	7	1.00
$1500 \leq \text{shadow size} < 1800$	5	5	1.00
$1800 \leq \text{shadow size} < 2100$	6	6	1.00
$2100 \leq \text{shadow size} < 2400$	3	3	1.00
$2400 \leq \text{shadow size} < 2700$	4	4	1.00
$2700 \leq \text{shadow size} < 3000$	1	1	1.00
$3000 \leq \text{shadow size} < 3300$	0	0	NA
$3300 \leq \text{shadow size} < 3600$	1	1	1.00
$3600 \leq \text{shadow size} < 3900$	0	0	NA
$3900 \leq \text{shadow size} < 4200$	1	1	1.00
$4200 \leq \text{shadow size} < 4500$	0	0	NA
$4500 \leq \text{shadow size} < 4800$	0	0	NA
$4800 \leq \text{shadow size} < 5100$	1	1	1.00
$5100 \leq \text{shadow size} < 5400$	1	1	1.00
$5400 \leq \text{shadow size} < 5700$	1	1	1.00
$5700 \leq \text{shadow size} < 6000$	0	0	NA
$6000 \leq \text{shadow size}$	1	1	1.00

Table 3.15 The sensitivity of detecting lesions by disease shadow size where the number of points is less than 300 using validation data from fold five.

Disease shadow size range	Number of detected lesions	Number of lesions	Sensitivity
$0 \leq \text{shadow size} < 10$	4	39	0.103
$10 \leq \text{shadow size} < 20$	20	36	0.556
$20 \leq \text{shadow size} < 30$	12	21	0.571
$30 \leq \text{shadow size} < 40$	13	22	0.591
$40 \leq \text{shadow size} < 50$	17	22	0.773
$50 \leq \text{shadow size} < 60$	16	22	0.727
$60 \leq \text{shadow size} < 70$	12	14	0.857
$70 \leq \text{shadow size} < 80$	15	20	0.75
$80 \leq \text{shadow size} < 90$	13	15	0.867
$90 \leq \text{shadow size} < 100$	6	8	0.75
$100 \leq \text{shadow size} < 110$	8	8	1.00
$110 \leq \text{shadow size} < 120$	8	8	1.00
$120 \leq \text{shadow size} < 130$	7	9	0.778
$130 \leq \text{shadow size} < 140$	7	8	0.875
$140 \leq \text{shadow size} < 150$	6	7	0.857
$150 \leq \text{shadow size} < 160$	5	5	1.00
$160 \leq \text{shadow size} < 170$	9	9	1.00
$170 \leq \text{shadow size} < 180$	3	3	1.00
$180 \leq \text{shadow size} < 190$	6	6	1.00
$190 \leq \text{shadow size} < 200$	6	6	1.00
$200 \leq \text{shadow size} < 210$	5	5	1.00
$210 \leq \text{shadow size} < 220$	2	2	1.00
$220 \leq \text{shadow size} < 230$	2	3	0.667
$230 \leq \text{shadow size} < 240$	5	5	1.00
$240 \leq \text{shadow size} < 250$	6	6	1.00
$250 \leq \text{shadow size} < 260$	1	1	1.00
$260 \leq \text{shadow size} < 270$	4	4	1.00
$270 \leq \text{shadow size} < 280$	1	1	1.00
$280 \leq \text{shadow size} < 290$	2	3	0.667
$290 \leq \text{shadow size} < 300$	1	1	1.00

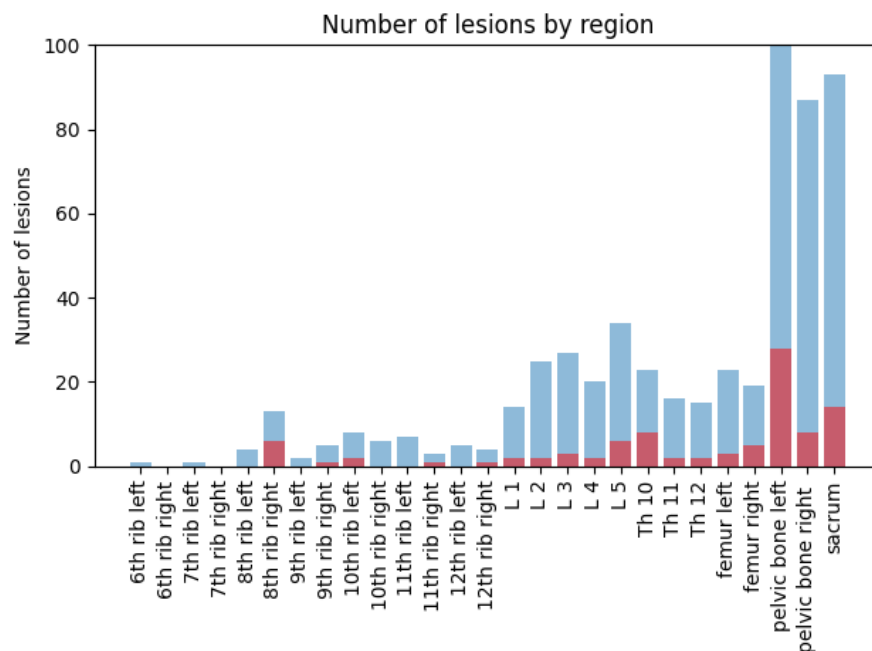


Fig. 3.24 The number of lesions grouped by region. The red bars show the number of missed lesions while the blue bars show the number of correctly detected lesions. This data is from the fold five validation data set.

lesions correctly (0.52). Table 3.17 shows the sensitivity of sclerotic lesions by shadow size and Table 3.16 shows the sensitivity of lytic lesion by shadow size.

An experiment was then carried out to verify the validity of labelling lesion types. This included selecting representative samples of the lesions and examining the results visually, as lytic lesions will contain a lesser bone density and sclerotic lesions will contain a greater bone density. Figure 3.25 shows an example of a lytic region being labelled correctly. In this figure the left shows the original CT image with the ground truth lesion shown in red. The right shows the same CT image with a white arrow pointing to the lesion. My technique labelled this lesion as lytic, which is correct as the bone density has decreased.

Smaller lesions will most likely be less accurately labelled, as there are fewer points to use in determining the relative symmetric bone density difference. And in cases where both sides are diseased, the method will probably be less accurate as well. This can be seen in Figure 3.26.

A check was carried out to see how many ground truth diseased shadow points were symmetrically mapped to other ground truth diseased shadow points. It was found that there was an overlap of 25% per case. This is significant, indicating that my lesion type labelling method was not accurate in certain cases. If the symmetric disease shadows from two regions

Table 3.16 The sensitivity of detecting lesions by disease shadow size for lytic lesions using validation data from fold five.

Disease shadow size range	Number of detected lesions	Number of lesions	Sensitivity
$0 \leq \text{shadow size} < 10$	1	21	0.048
$10 \leq \text{shadow size} < 20$	4	16	0.250
$20 \leq \text{shadow size} < 30$	3	11	0.273
$30 \leq \text{shadow size} < 40$	3	8	0.375
$40 \leq \text{shadow size} < 50$	7	11	0.636
$50 \leq \text{shadow size} < 60$	4	9	0.444
$60 \leq \text{shadow size} < 70$	4	6	0.667
$70 \leq \text{shadow size} < 80$	5	7	0.714
$80 \leq \text{shadow size} < 90$	4	5	0.800
$90 \leq \text{shadow size} < 100$	3	4	0.750
$100 \leq \text{shadow size} < 110$	3	3	1.00
$110 \leq \text{shadow size} < 120$	3	3	1.00
$120 \leq \text{shadow size} < 130$	1	3	0.333
$130 \leq \text{shadow size} < 140$	2	3	0.667
$140 \leq \text{shadow size} < 150$	1	2	0.50
$150 \leq \text{shadow size} < 160$	1	1	1.00
$160 \leq \text{shadow size} < 170$	1	1	1.00
$170 \leq \text{shadow size} < 180$	1	1	1.00
$180 \leq \text{shadow size} < 190$	1	1	1.00
$190 \leq \text{shadow size} < 200$	2	2	1.00
$200 \leq \text{shadow size} < 210$	1	1	1.00
$210 \leq \text{shadow size} < 220$	0	0	NA
$220 \leq \text{shadow size} < 230$	0	0	NA
$230 \leq \text{shadow size} < 240$	1	1	1.00
$240 \leq \text{shadow size} < 250$	0	0	NA
$250 \leq \text{shadow size} < 260$	0	0	NA
$260 \leq \text{shadow size} < 270$	1	1	1.00
$270 \leq \text{shadow size} < 280$	0	0	NA
$280 \leq \text{shadow size} < 290$	0	1	0.00
$290 \leq \text{shadow size} < 300$	0	0	NA
$300 \leq \text{shadow size}$	15	15	1.00

Table 3.17 The sensitivity of detecting lesions by disease shadow size for sclerotic lesions using validation data from fold five.

Disease shadow size range	Number of detected lesions	Number of lesions	Sensitivity
$0 \leq \text{shadow size} < 10$	3	18	0.167
$10 \leq \text{shadow size} < 20$	16	20	0.80
$20 \leq \text{shadow size} < 30$	9	10	0.90
$30 \leq \text{shadow size} < 40$	10	14	0.714
$40 \leq \text{shadow size} < 50$	10	11	0.909
$50 \leq \text{shadow size} < 60$	12	13	0.923
$60 \leq \text{shadow size} < 70$	8	8	1.00
$70 \leq \text{shadow size} < 80$	10	13	0.769
$80 \leq \text{shadow size} < 90$	9	10	0.90
$90 \leq \text{shadow size} < 100$	3	4	0.75
$100 \leq \text{shadow size} < 110$	5	5	1.00
$110 \leq \text{shadow size} < 120$	5	5	1.00
$120 \leq \text{shadow size} < 130$	6	6	1.00
$130 \leq \text{shadow size} < 140$	5	5	1.00
$140 \leq \text{shadow size} < 150$	5	5	1.00
$150 \leq \text{shadow size} < 160$	4	4	1.00
$160 \leq \text{shadow size} < 170$	8	8	1.00
$170 \leq \text{shadow size} < 180$	2	2	1.00
$180 \leq \text{shadow size} < 190$	5	5	1.00
$190 \leq \text{shadow size} < 200$	4	4	1.00
$200 \leq \text{shadow size} < 210$	4	4	1.00
$210 \leq \text{shadow size} < 220$	2	2	1.00
$220 \leq \text{shadow size} < 230$	2	3	0.667
$230 \leq \text{shadow size} < 240$	4	4	1.00
$240 \leq \text{shadow size} < 250$	6	6	1.00
$250 \leq \text{shadow size} < 260$	1	1	1.00
$260 \leq \text{shadow size} < 270$	3	3	1.00
$270 \leq \text{shadow size} < 280$	1	1	1.00
$280 \leq \text{shadow size} < 290$	2	2	1.00
$290 \leq \text{shadow size} < 300$	1	1	1.00
$300 \leq \text{shadow size}$	76	76	1.00

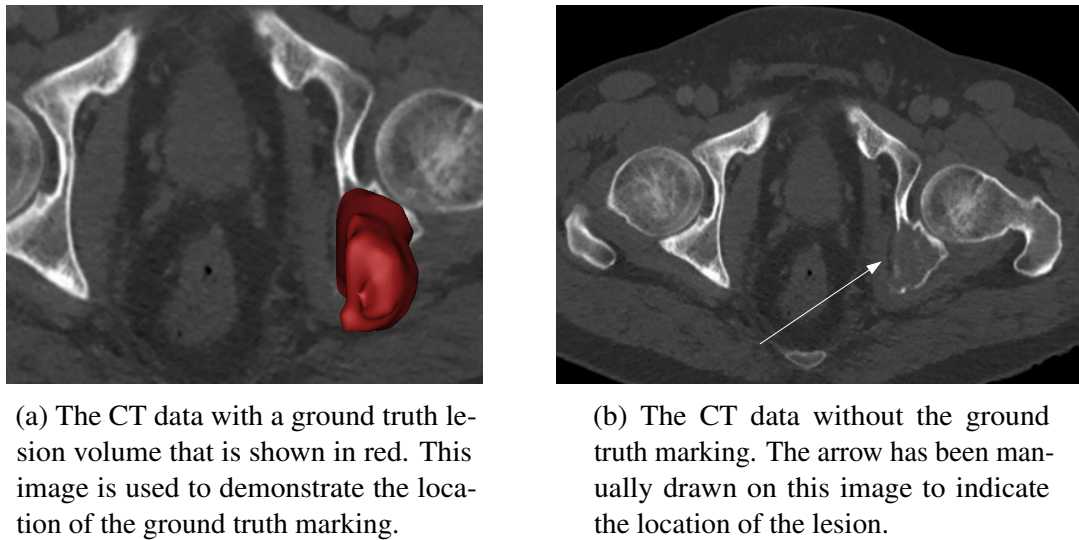


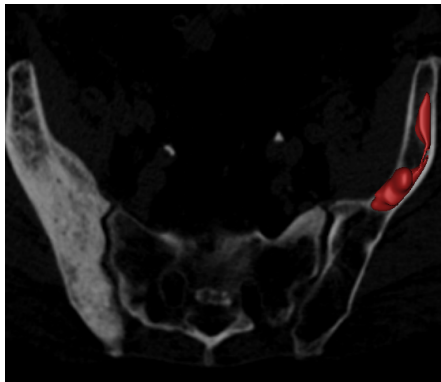
Fig. 3.25 An example of correctly labelling a lesion lytic. My algorithm labelled this lesion by symmetrically comparing the CBM trabecular bone density measurements. The CT data is from the Addenbrooke's Hospital CT Data Collection.

fully overlap, one side will be labelled as lytic and the other as sclerotic. This could cause inaccurate labelling, as both sides could actually be sclerotic (the sclerotic lesion with less bone density between the two would be labelled as lytic) or vice versa. If a partial overlap is present, the lesions may both still be labelled correctly as there is healthy bone to compare with. However, it appears that this method did label most lesions correctly although this was not verified by a radiologist.

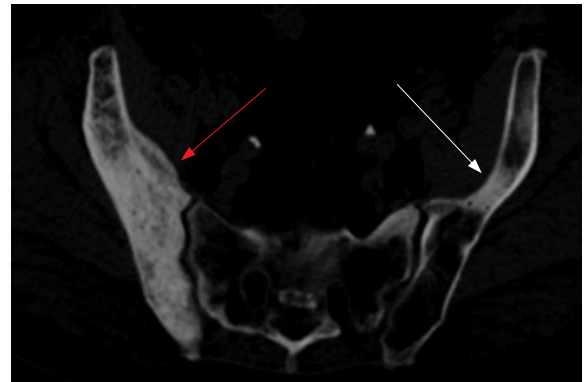
3.4.10 Diseased side selection

The results from Section 3.4.8 used the diseased side selection method described in Section 2.5.3 to determine whether a region was diseased or healthy. It was found that 4.01% of all points that were incorrectly marked as diseased had their symmetrical mapped point labelled incorrectly as being healthy (as it was diseased) using validation data from fold five. A discussion about these results are located in Section 4.2.6.

As the majority of the lesions in this data set are sclerotic (having an increase in trabecular bone density), another experiment was carried out where the side selection method (section 2.5.3) was disregarded. For this experiment, it was assumed that every diseased region was sclerotic. Instead of comparing border gradients of the trabecular bone density, the trabecular bone density was compared directly. For each point in a region, the trabecular bone density was summed. The symmetrically mapped point was also looked up for each point in the region and the trabecular bone density at each of these symmetrical mapped



(a) The CT data with a ground truth lesion volume that is shown in red. This image is used to demonstrate the location of the ground truth marking. The left side is also diseased although the ground truth visualisation is not displayed.



(b) The CT data without the ground truth marking. The white arrow has been manually drawn on this image to indicate the location of the lesion. The red arrow shows where the symmetrically mapped region is likely to be located.

Fig. 3.26 A example of incorrectly labelling a lesion as lytic. My algorithm labelled this lesion as lytic by symmetrically comparing the CBM trabecular bone density measurements. However, this lesion is actually sclerotic (although this has not been verified by a radiologist). This lesion was most likely labelled incorrectly as its symmetrically mapped points are also diseased and sclerotic as seen by the denser white bone on the left side of the image. The CT data is from the Addenbrooke's Hospital CT Data Collection.

Table 3.18 The table shows the number of detected and missed lesions out of 1626 using line data sampled at 40 locations using 2 smoothing iterations and a cluster size of 7. The full data set was used in the experiment (instead of using a cross validation approach). Results would most likely be worse using CV approach.

HU threshold	point accuracy	average sensitivity
100	0.927	0.541
1	0.882	0.805
5	0.886	0.799
4	0.885	0.801
2	0.883	0.801
3	0.884	0.801
50	0.913	0.677
10	0.891	0.778
20	0.898	0.747
45	0.912	0.690

points was also summed. The region with the higher total was marked as diseased. This method detected 967 out of 1067 sclerotic lesions correctly producing a sensitivity of 0.906 using the parameters optimised in Section 3.4.8.

This experiment was also performed for lytic lesions by assuming all lesions were lytic (having a decrease in trabecular bone density) and by marking a region as diseased if the first total had a smaller sum than the second. This method correctly detected 427 out of 501 lytic lesions producing a sensitivity of 0.852.

3.4.11 Experiments using line data

Instead of using CBM measurements for comparison, the raw line data was used directly for detection. In this experiment, at a number of evenly spaced increments along the 10.8 mm line normal to the bone, radiodensity measurements were taken using the Stradview software [133]. For these experiments, 40 measurements were taken along each 10.8 mm line segment providing a radiodensity sample every 0.27 mm ($10.8/40 = 0.27$ mm). As the minimum x and y pixel spacing is 0.580 mm and minimum slice spacing 2 mm, sampling every 0.27 mm was a short enough distance to accurately measure the data.

The measurements at each vertex were compared between symmetric points by taking the absolute values of their summed differences. This value was compared to a threshold and points above the threshold were marked as diseased. A range of thresholds were used and the results from this experiment are listed in Table 3.18. The results from this table produce significantly lower sensitivities and point accuracies.

3.4.12 Experiments on atlas symmetrical deformation mappings

Instead of using the segmentation mesh for comparison, in this experiment the atlas is deformed and used to match the skeleton as described in Section 2.4.5. This method produced a sensitivity of 0.846 and a point accuracy of 0.717 using the parameters optimised using cross validation in Section 3.4.8. This method detected 1376 lesions out of 1626 correctly. The point accuracy is significantly worse than using the segmentation directly as described above.

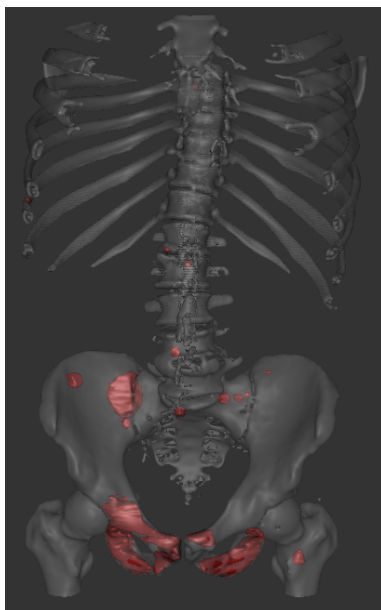
3.5 Longitudinal evaluation

In longitudinal assessment, baseline and follow-up cases were compared with each other as described in Section 2.6. The ground truth for the longitudinal results was determined by starting with the ground truth disease shadows (as described in Section 3.2.3) for each case. The registration point mapping data was used to compare the label of each point on the baseline segmentation to its mapped point in the follow-up case. For the points in which the label changed from healthy to diseased, the point was marked as a disease progression point. A point was labelled as disease regression if it changed from diseased to healthy.

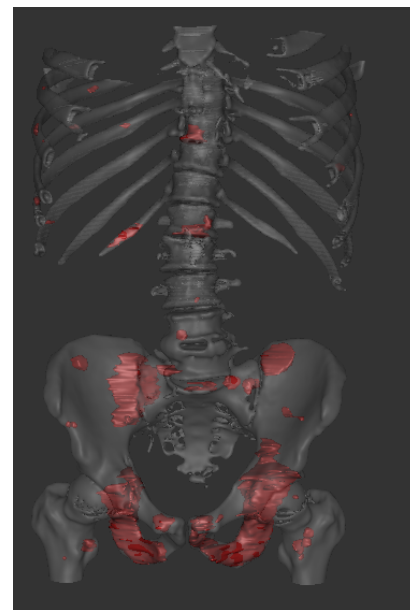
For assessment, the trabecular bone density of each point in the base case segmentation was compared with the trabecular bone density of each point in the follow-up segmentation using the registration point mapping. Differences above a defined threshold were labelled as diseased. Connected groups smaller than a defined cluster size were discarded as described in Section 3.4.3. The results from the longitudinal cases are shown in Table 3.19 using the parameters optimised using the cross validation parameters from Section 3.4.8. Table 3.20 shows the results where the HU threshold and the cluster size were allowed to vary (without cross validation). A discussion about longitudinal analysis is located in Section 4.3.2.

Figure 3.27 shows an example of baseline and follow-up cases for a single patient. It can be seen that the disease has progressed from the baseline case to the follow-up case as the lesions have increased in size. New lesions have also appeared within the follow-up case that were not present in the baseline case. The longitudinal assessment results for this case are shown in Figure 3.28. Figure 3.29 shows another baseline and follow-up case. Figure 3.30 shows the results for this case which are poorer than the previous case.

In another experiment, the longitudinal assessment method was used to predict lesion volume growth. This was calculated by taking the longitudinal diseased shadow size (spread of disease) and by using this number with equation (3.3) (to find lesion volume). In order to assess this result, a ground truth volume was compared to the longitudinal assessment's predicted volume growth.



(a) A baseline case – an alpha blend is used here to show the 3D lesions (shown in red) within the bony surface.



(b) The follow-up case – it can be seen that the 3D lesions have primarily increased in size from the baseline case.

Fig. 3.27 A typical example of the disease progression. The bony segmentations have been displayed in grey while the lesions have been displayed in red. The disease has progressed from the baseline case to the follow-up case.

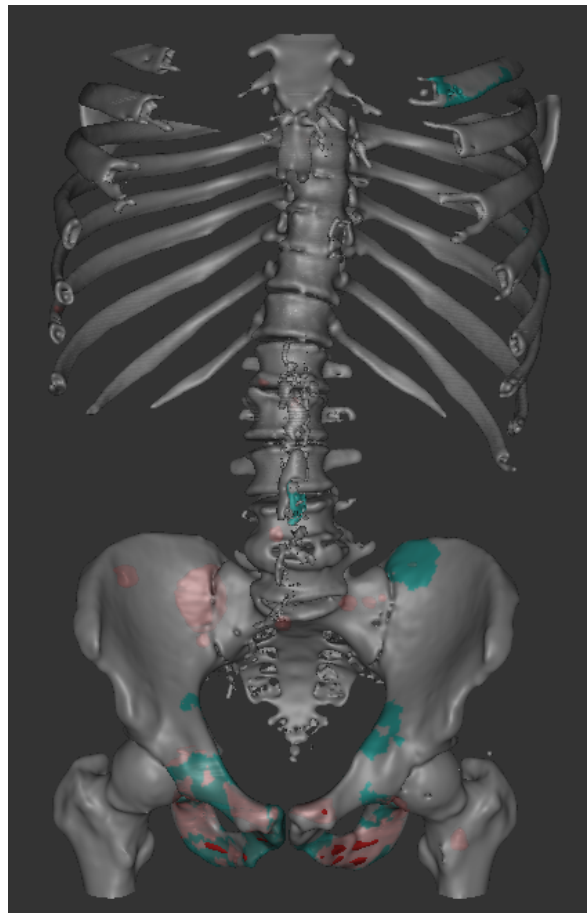
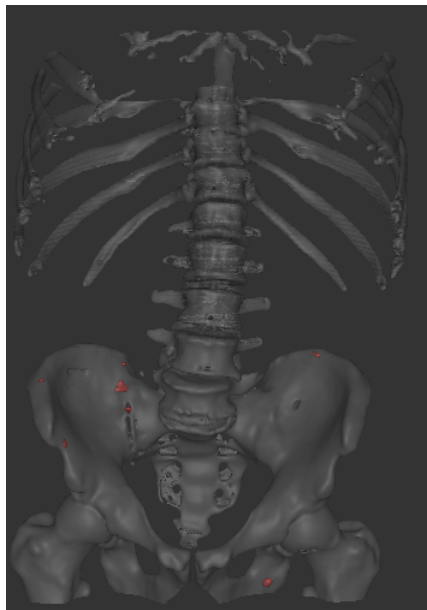
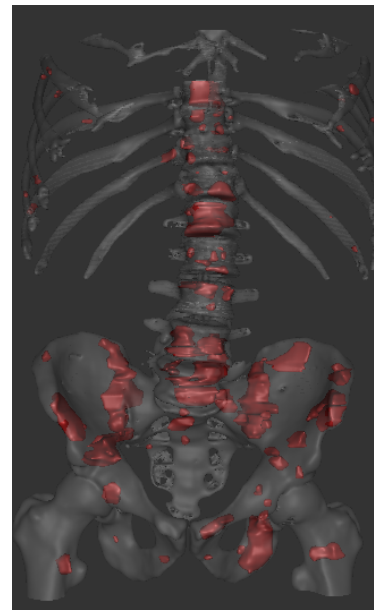


Fig. 3.28 An example of the results from a longitudinal experiment. In the longitudinal assessment, strong differences in trabecular bone density between baseline and follow-up cases are marked as disease. These diseased regions are shown in teal. The baseline case had a total sum ground truth lesion volume of 120.845 cm^3 . The follow-up case had a total sum ground truth lesion volume of 264.592 cm^3 . This was an increase of 143.747 cm^3 . The predicted volume for this case was 101.442 cm^3 using the longitudinal assessment method. This is an absolute percent error difference of 29.431%



(a) A baseline case, the 3D lesions are shown in red.



(b) The follow-up case contains many more lesions.

Fig. 3.29 An example of longitudinal disease progression. It can be seen that the 3D lesions have significantly increased in number and in size from the baseline case to the follow-up case.

Table 3.19 The point accuracy using the optimised parameters described in 3.4.8 (using the parameters optimised by cross validation on the single data sets).

Accuracy	HU threshold	Smoothing iterations	Cluster size
0.807	150	2	7
0.773	125	2	7
0.723	100	2	7

To find the ground truth volume growth, the total lesion volume was found for each case (by summing each lesion's volume for every lesion within a case). Then, the absolute value of the volume differences between each baseline and follow-up cases was found to determine the ground truth. In Table 3.21, the ground truth volume growth has been compared to the longitudinal assessment's predicted volume growth. This comparison shows the percentage difference between the two values. Table 3.22 shows 10 representative cases, which include the actual baseline and follow-up volumes. It also shows the growth and estimated growth that was found by converting diseased shadow points to volume using equation (3.3).

Table 3.20 The 10 best longitudinal accuracies letting the HU threshold and the cluster size vary. Ranges were set near to the optimised parameters described in 3.4.7 (using cross validation to optimise the parameters on single data sets), although cross validation was not used to optimise these results (across multiple data sets).

Accuracy	HU threshold	Smoothing iterations	Cluster size
0.925	150	2	600
0.918	150	2	400
0.915	125	2	600
0.911	150	2	200
0.910	125	2	400
0.908	150	2	150
0.906	150	2	125
0.903	150	2	100
0.901	125	2	200
0.898	100	2	600

Table 3.21 The 10 best results by volume percentage difference. The volume percentage difference is the absolute value of the calculated volume subtracted from the ground truth volume divided by the ground truth volume. Initial ranges were set near to the optimised parameters described in 3.4.7 (using cross validation to optimise the parameters on single data sets), although cross validation was not used to optimise these results (across multiple data sets).

Volume percentage difference	HU threshold	Smoothing iterations	Cluster size
107.41 %	150	2	400
111.78 %	150	2	600
140.56 %	150	2	200
153.39 %	150	2	150
154.81 %	150	2	125
161.43 %	150	2	100
169.41 %	125	2	600
174.40 %	125	2	400
192.10 %	150	2	50
209.25 %	125	2	200

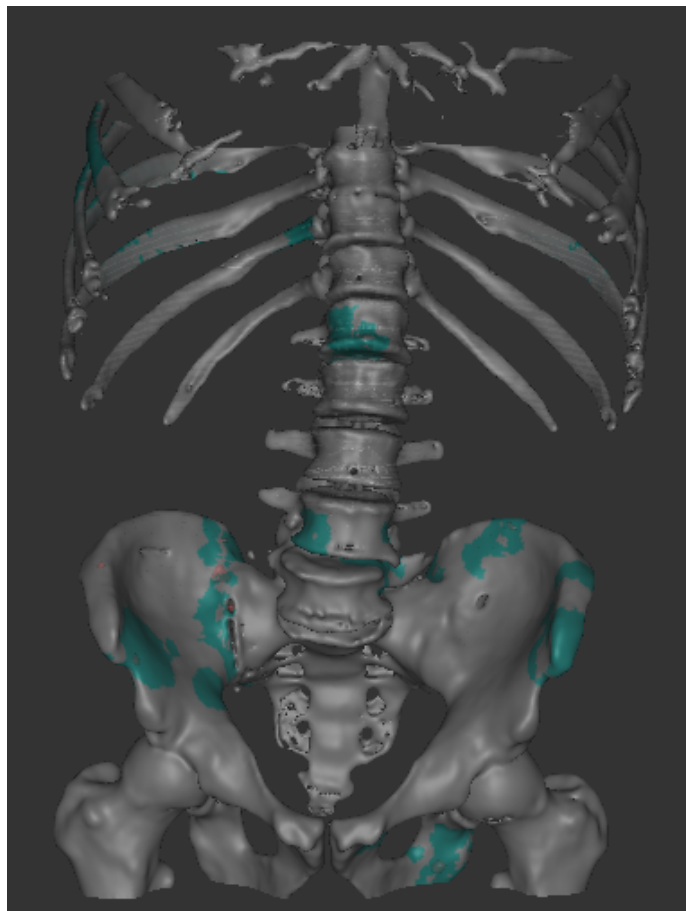


Fig. 3.30 An example of a longitudinal result. The change in disease between baseline and follow-up cases is shown in teal on the bony surface. The baseline case had a total sum ground truth lesion volume of 1.249 cm^3 . The follow-up case had a total sum ground truth lesion volume of 305.602 cm^3 . This was an increase in 304.352 cm^3 . The predicted volume for this case was 150.501 cm^3 using the longitudinal assessment method. This is an absolute percent error difference of 50.6%.

Table 3.22 10 representative longitudinal cases showing total lesion volume (cm^3) of the baseline and the follow-up scans. Volume growth is the change in volume (cm^3) from baseline to follow-up scans. The volume estimate (cm^3) is the predicted volume growth which uses equation (3.3) to convert shadow points to volume. Percentage difference shows the relationship between the volume growth and the volume estimated growth.

Case	Baseline	Follow-up	Growth	Estimate	% difference
1	0.86	59.94	59.07	131.57	122.72%
2	191.80	276.73	84.93	11.79	86.12%
3	1.32	0.20	-1.12	0.00	100.00%
4	42.41	223.01	180.60	10.63	94.12%
5	120.85	264.60	143.75	94.65	34.16%
6	1.25	305.60	304.35	150.50	50.55%
7	12.18	5.50	-6.68	4.53	167.83%
8	318.82	416.89	98.07	65.47	33.24%
9	34.87	264.84	229.97	65.50	71.52%
10	80.26	1752.66	1672.40	877.67	47.52%

Table 3.23 The table shows running times for the different executables in the symmetrical assessment method.

Name	Min	Max	Average
Segmentation	3m 49s	12m 11s	8m 45s
CBM and bone mesh generation	2m 35s	51m 51s	16m 16s
MBD mesh generation	1s	30m 25s	3m 23s
Registration	17m 59s	1h 21m 55s	34m 12s

3.6 Running times

The running times for the executables are listed in Table 3.23. These jobs were run on an Alienware Aurora R8 gaming desktop (9th Gen Intel(R) Core(TM) i7 9700 (8-Core), 32GB DDR4 and NVIDIA GeForce GTX 1660Ti with 6GB GDDR6) running the Ubuntu operating system. All executables were single threaded, although batches of jobs were run at the same time across the multiple cores. The data passed through a pipeline where the output from each executable becomes the input for the next. A diagram of this execution pipeline can be seen in Figure 3.31.

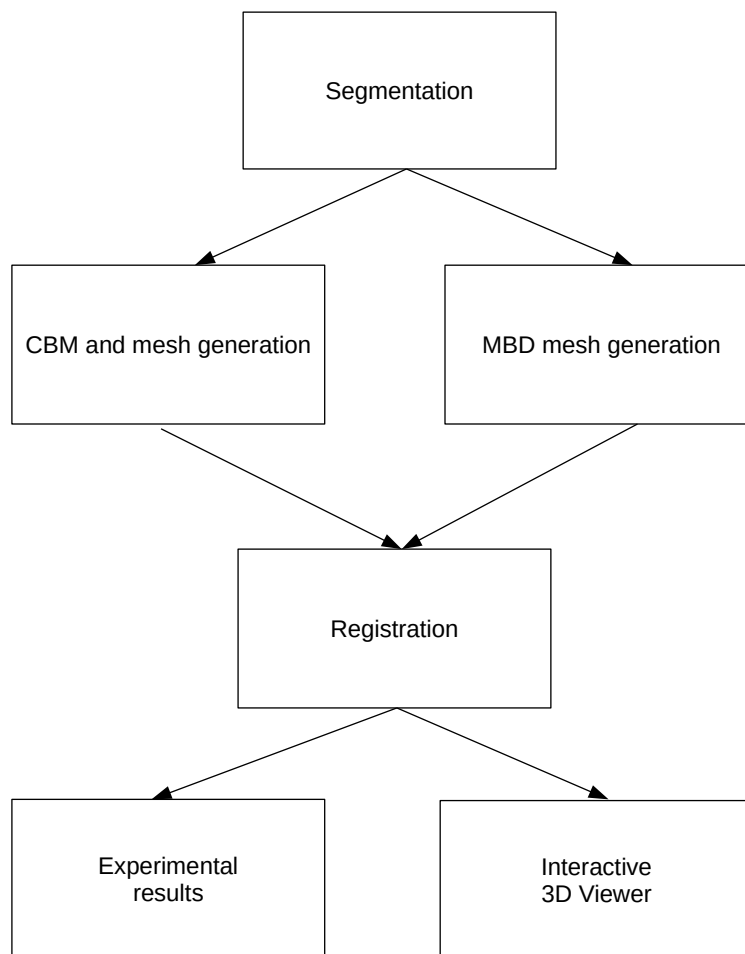


Fig. 3.31 This image shows the flow of the execution pipeline where each box is an executable. The output of each executable is used as the input of the next. The CBM and mesh bone generation executable and the MBD mesh generation executable can be run in parallel. Results can be calculated using the experimental results executable as well as viewed interactively in the 3D viewer or both.

Table 3.24 The table shows running times for the different executables for the longitudinal assessment.

Name	Min	Max	Average
Segmentation	8m 21s	13m 30s	10m 00s
CBM and bone mesh generation	1m 28s	27m 32s	10m 59s
MBD mesh generation	5s	24m 32s	5m 18s
Registration	16m 39s	51m 48s	35m 09s

Chapter 4

Discussion

4.1 Introduction

In this chapter, I will be addressing a number of points regarding the algorithms, including the input data sets, segmentation, registration, overall results, usefulness of the approach as a whole and the various limitations.

4.2 Experimental data

Using experimental data that is representative of the actual diseased data (as would be encountered in a clinical setting) is essential in demonstrating the validity of the symmetrical algorithm and the longitudinal algorithm. As all of the diseased data used in these experiments was produced by imaging real prostate cancer patients, the experimental data is a valid representation of real clinical data. Because of this, it is likely that the results produced by my algorithm, in the experiments described in this thesis, would be similar to the results produced in a clinical setting.

However, having a high number of data sets from a wide range of patients is also important, since the results can only be considered reliable on patients with the same demographics, and the same range of disease, as is contained in this data. As the data sets used in this thesis overlap the expected age range and disease progression of a patient with prostate cancer, these data sets can successfully provide an accurate indication of how this method will perform on real prostate cancer patients. Having many CT data sets from a range of differing bone anatomies, shape, sizes and ages also reduces the chance of the algorithm over-fitting to the specific experimental data. The number of CT data sets used in these experiments was 103, with 68 being diseased and 35 being healthy. This is similar to the

average number of data sets used by other researchers in their experiments in detecting MBD, so these tests are expected to be as robust as those typically reported in the related literature.

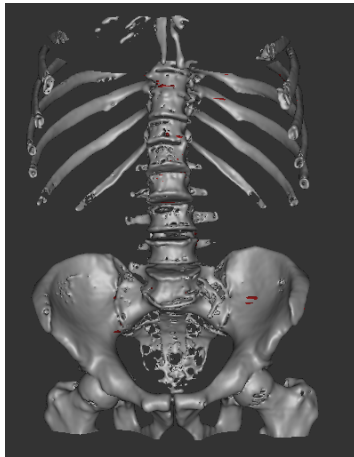
Although 103 data sets is a significant number, more data would be beneficial as it would provide a more robust indication of the true sensitivity and accuracy produced by this algorithm. By using a five fold cross validation technique to optimise the three parameters (as described in Section 3.4.6), over-fitting of the data can be reduced. However, a limitation of this approach is that data was not available for a testing set. A test set would have provided a more unbiased evaluation of the results.

Unfortunately, it is very difficult to obtain these CT scans due to the usual issues surrounding clinical ethics in research, as well as the practical issues including patient consent forms, anonymisation, release and the secure transfer of the data, even though it is common for prostate cancer patients to be scanned in a CT scanner. In the project, attaining usable CT data took a significant amount of time (well over a year) and expertise. Once these were acquired, Dr. Barrett spent a considerable period of time to mark the outlines of the lesions. This acquisition process is challenging to do on a large scale due to the time constraints, especially for a busy radiologist, further limiting the number of data sets that are practically available.

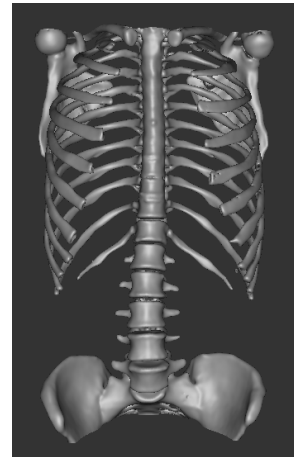
The lesion type labels are missing from the marked data, which is important. However, it was not possible to obtain these due to the radiologist's time constraints and due to the outbreak of the pandemic. However, an alternative method for determining the lesion type was developed as described in Section 3.4.9. These results have not been verified by the radiologist, but upon visual inspection they appear to be largely accurate.

With regards to disease, the majority of the lesions in these data sets have a volume of less than 3.0 cm^3 and are generally located around the pelvis. The locations of the lesions are determined by the primary cancerous tumour cells that travel through a patient's arteries and veins [100].

There are also slight discrepancies between the diseased data sets and the healthy data sets. The diseased data sets extend lower than the healthy data sets, and usually contain the upper femurs but do not contain the upper ribs as seen in Figure 4.1a. The healthy data sets extend higher than the diseased data sets, and include the shoulders and upper ribs but often do not contain the upper femurs. This can be seen in Figure 4.1b. It would be beneficial if both the healthy and diseased CT images overlapped in exactly the same regions, but this is constrained by the scanning protocol. This was not controlled by this study; the healthy data represented an opportunistic sample. This is because the European Association of Urology (EAU) only recommends abdomen and pelvis coverage at baseline staging. The majority of



(a) An example of a typical diseased data set that extends from the ribs to the lower femurs.



(b) An example of a healthy data set that extends from the shoulders to mid pelvis.

Fig. 4.1 Examples of typical diseased and healthy data sets. The healthy data sets extend to the shoulders, while the diseased data sets extend to the upper femurs.

the bony regions do overlap, however this also shows that the algorithm can work correctly on slightly different regions of the skeleton.

4.2.1 Segmentation results

Automatic segmentation of bone in CT is an essential part of the detection of metastatic bone disease. Creating perfect segmentations is, however, a challenging problem. Much of the difficulty is caused by the noisy CT data with often low contrast values between the bone and non-bone. In many cases, it is not even clear how to correctly label bone during a manual segmentation process. Errors in the segmentation are compounded as segmentation precedes the other steps in the algorithm. Poor results in the segmentation can cause poor results in the registration process, and can lead to a reduction in performance of the symmetry method's overall sensitivity and point accuracy.

My segmentation method used in these experiments produced an average DSC score of 98.95%, as shown in Section 3.3.1. This is more accurate than the other segmentation techniques described in section 2.2.1, which produced DSC scores of (98.20%, 95.00%, 90.00% and 84.40%). Significantly, my method outperforms Fu et al. [43] by 8.85% even though it is based on the same core segmentation technique (hysteresis thresholding) and despite that fact, unlike their version, my method does not require any human interaction.

The total average surface distance produced for the segmentations using my implementation of the algorithm was 2.168 mm. This is slightly worse than the 0.41 mm - 1.55 mm

produced by Fu et al. [43], although the segmentation was still good enough to produce usable results in both the symmetrical and longitudinal algorithms. Fu et al. [43] may have achieved a lower average surface distance error because it is not a fully automatic method but requires some human intervention.

My method produced a better DSC score than Fu et al. [43], but also had a greater average surface distance error than Fu et al. [43]. This means that my method had a higher accuracy in labelling all voxels correctly but was less accurate in generating a correct bony surface. As there are many more non-bone voxels than bony voxels, my segmentation method is likely more accurate in labelling non-bone voxels correctly than in labelling bony voxels correctly.

Related to this, my segmentation algorithm achieved a precision of 0.92. This is the accuracy limited to voxels that have been labelled as bone. This result was similar to 0.94, which was achieved by [71], and less than 0.99, which was achieved by [76]. These methods are likely to produce a more accurate labelling of bone than my method. The precision achieved by Fu et al. [43] method was not available, but it was most likely greater than my precision of 0.92 as Fu et al. [43] achieved a smaller average surface distance error.

As highlighted in Figure 4.2, the segmentation errors are common in a number of areas. This most often occurs when a bony voxel's radiodensity value overlaps the normal range for the radiodensity of either bone or tissue. Both tissue and bone can be binned into the same voxel during the image reconstruction process, which makes it difficult to distinguish between the two.

My segmentation algorithm has difficulties with the vertebrae as shown in Figure 4.3, as they are especially hard to segment correctly. The 2D slices often cut through the patient in each vertebra at an angle that is challenging to reconstruct back into the 3D model. Figure 4.4 shows a segmentation that produced a far worse result than normal, which is an example of where this issue was present.

However, the overall segmentation results are acceptable for use in this algorithm. This can be clearly seen in Section 3.4.8, as an overall sensitivity of 0.820 was achieved in detecting both sclerotic and lytic lesions within the same data set.

My segmentation method took an average of 10 minutes per case. This is significantly faster than Fu et al. [43] (who used a similar segmentation technique), as this method took about 20 minutes to run per skeleton. The machine learning methods are significantly faster – an average of 8.13 minutes using the random forest method [14], less than 2 or 3 minutes using the watershed and support vector machine method [76], and under one second using the U-net method [71].

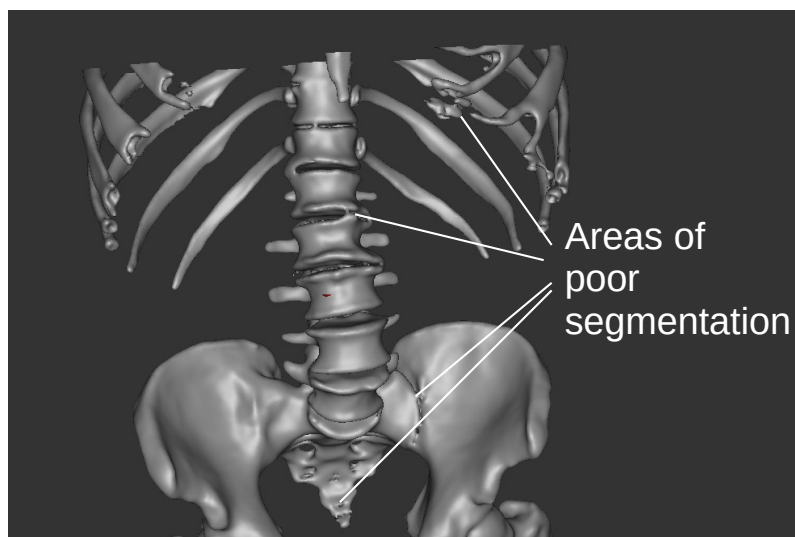


Fig. 4.2 A typical segmentation – the segmentation is not accurate in places. This is due to the partial volume effect, as both bone and non-bone is binned into the same voxel.

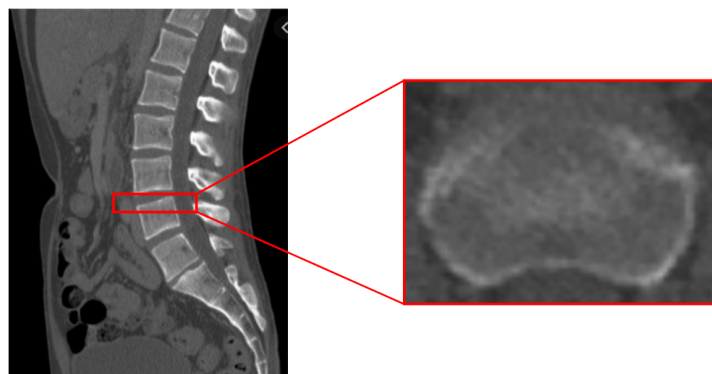


Fig. 4.3 An example of a difficult region to accurately segment. Segmentation of the vertebrae is challenging due to the partial volume effect and the slice angle through each vertebra. An example of a typical vertebra in a CT scan – it is unclear where the upper segmentation boundary should be in this case due to the small angle difference between the surface of the vertebra and the CT slice. The CT data is from the Addenbrooke's Hospital CT Data Collection.

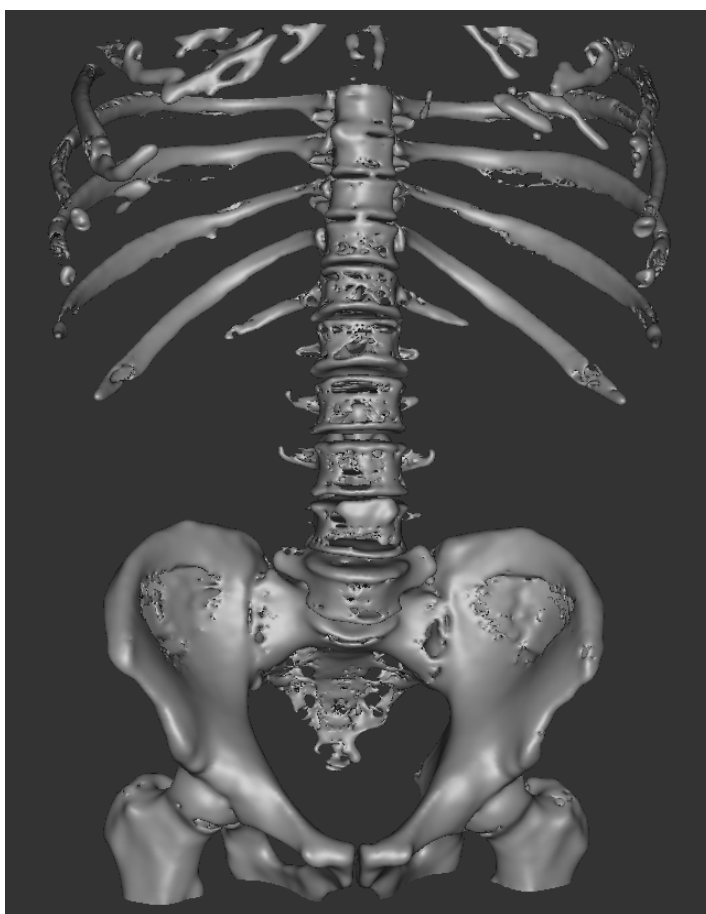


Fig. 4.4 A typical poor quality segmentation result – the vertebrae are particularly problematic due to the extremely thin cortices.

4.2.2 Registration results

Registration is an essential process in both the symmetrical and the longitudinal methods. It either determines the symmetrical mapping between points (in the case of the symmetrical method) or the regional mapping between baseline and follow-up points. Errors in registration will cause the trabecular bone density to be compared at incorrect points, which can increase the overall error in both algorithms.

Generally, the majority of the registration is very accurate. Figure 4.5 shows what a typical registration result looks like. In this image, the segmentation mesh is slightly interleaved with the deformed atlas in many places. This indicates the registration matches are very close to each other in this situation.

The overall average distance error between the registered atlas and segmentation was 4.84 mm. As seen in Table 3.5, most of the larger errors are due to problems in registering the ribs as they are much more difficult to register correctly. The pelvic bones contained a much smaller error of 1.78 mm, 1.87 mm and 2.70 mm (for the pelvic bone left, pelvic bone right and sacrum respectively). This is significant as a majority of the lesions (as shown by Figure 3.4), are located within the pelvic bones. Only a low number of lesions are located within the ribs.

Figure 3.4 shows that another common location of the lesions is in specific vertebrae and in the femurs. Most of the vertebrae have an average distance error significantly lower than the overall distance error of 4.84 mm. When excluding vertebra Th 10, these values ranged from 1.91 mm to 2.66 mm. Th 10 has an average distance error of 8.14 mm, that is greater than the other vertebrae as it often does not have a segmentation vertebra to match to, as the atlas extends higher than the segmentation.

There is, however, a noticeable amount of error within the registration results in certain cases. This registration error can be due to a variety of issues, including errors in the segmentation, asymmetric bone, large differences in shape and size between the segmentation and atlas, difficult original registration placement issues and imperfect point matching criteria. Figure 4.6 shows a region where a large difference between shape and size of the atlas and segmentation was significant enough to produce poor registration matching on the edge of the pelvis. Figure 4.7 shows the results of a poor point symmetry match.

The largest error in the registration is in the ribs. These errors are primarily caused by the large differences in shape, size, length as well as the difference in number of ribs between the segmentation and the atlas. These errors do not significantly affect the sensitivity, as most of the lesions are in the spine, pelvis and femurs. However, they can reduce the overall point accuracy of the algorithm.

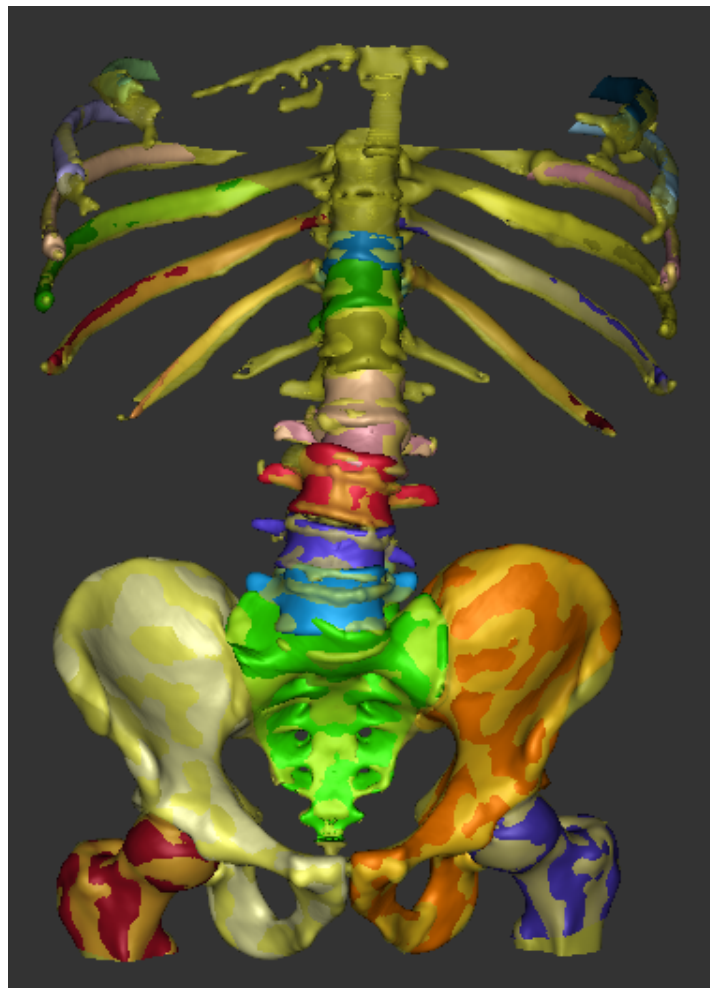


Fig. 4.5 A typical deformable registration result. The segmentation mesh is displayed in yellow and the individual pieces of the atlas are displayed by specific colours. The registration results are especially accurate within the pelvis and the femurs where the majority of the disease is located. It is important to note that there are some issues with the vertebrae registration.

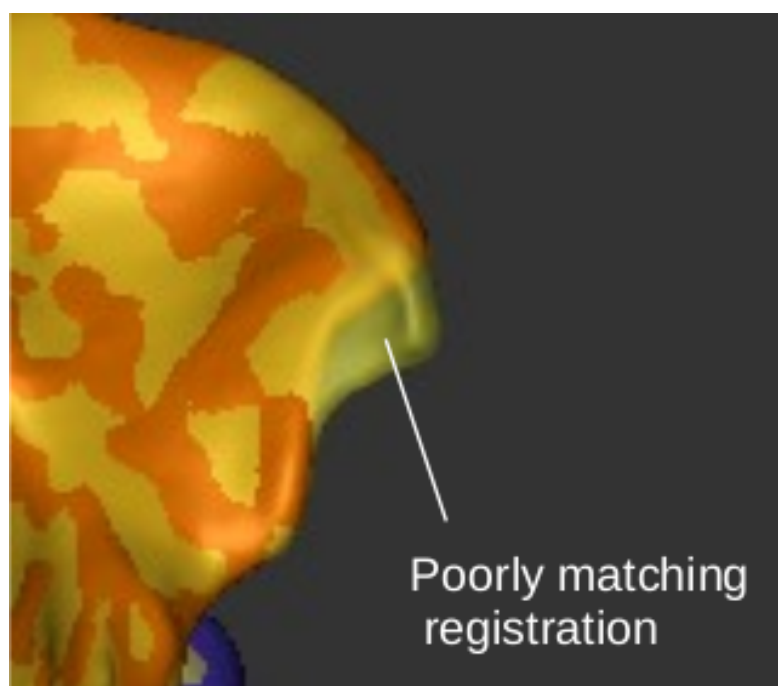


Fig. 4.6 The registration is not accurate in places as shown in this image. Significant differences in shape and size of the atlas and the segmentation can cause areas of poorly matching registrations.

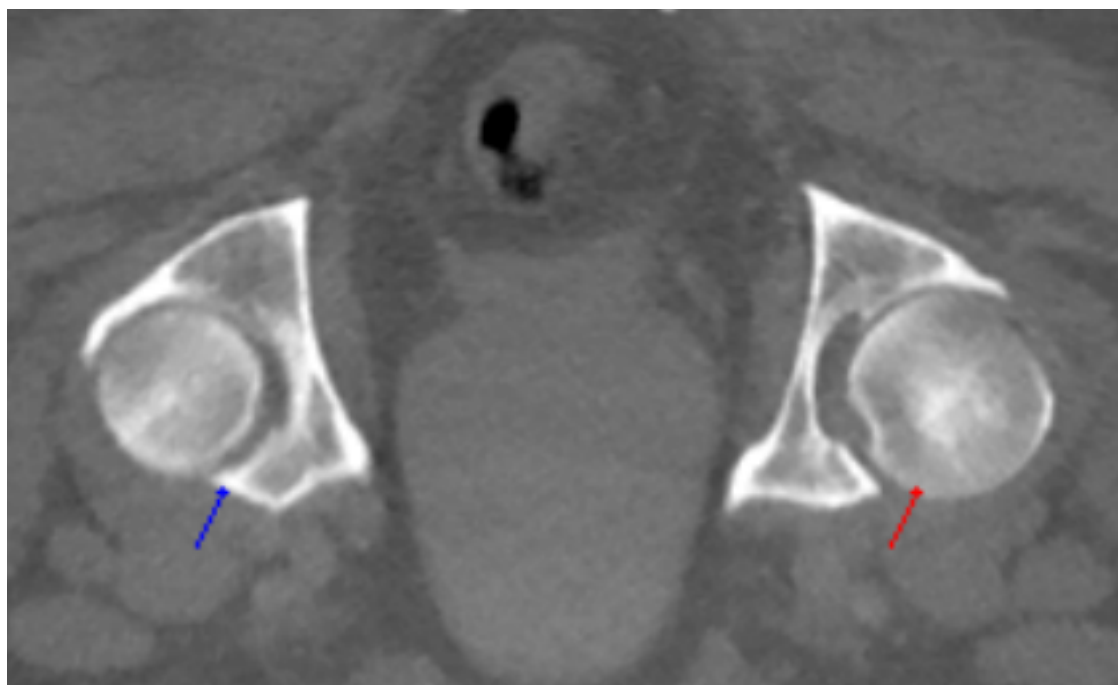


Fig. 4.7 An example of a poor symmetrical point match. This will cause a decrease in the overall accuracy of the algorithm, since clearly the wrong points are being compared. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.

Overall, there are errors in the registration from a number of sources that negatively affect the final results of the algorithm. While much of the registration is correct and there is still enough symmetry for the symmetrical method to produce good results, improved registrations would certainly produce better results.

4.2.3 Parameter choice

Setting the correct parameters is essential to producing good results for the symmetrical and longitudinal methods. When determining the parameters used by the algorithm, there are a number of factors to consider. The first of these is a minimum number of smoothing iterations and a minimum cluster size needed to remove the speckle noise that exists when using these methods. Removing speckle noise can be achieved using at least 2 smoothing iterations or a minimum cluster size of 7, as described in section 3.4.5. Without removing the speckle noise, the results are effectively invalid as the speckle noise will incorrectly detect all lesions simply because the noise is generally spread relatively evenly across all of the bony mesh.

Another significant determining factor in setting the parameters is the size of the lesions. When projecting symmetric difference in trabecular bone density to the surface, lesions produce a disease shadow as described in Section 3.2.3. As shown in Figure 3.6, there is approximately a power of 2 to 3 relationship between number of points in a disease shadow to the total volume in a lesion. This relationship occurs because the number of points is a measure of squared millimetres while volume is a measure of cubed millimetres, although it is only approximate since the lesions can have a very complex shape.

Larger lesions produce larger disease shadows. A larger disease shadow can be detected using a larger cluster size parameter. Using a larger cluster size parameter improves the overall point accuracy as much of the noise can be filtered away. Therefore, the largest possible cluster size should be used in order to detect the lesions. Figure 4.8 shows a large legion producing a large disease shadow. The cluster size can be set to a very high number and still successfully detect this legion.

Smaller lesions produce smaller disease shadows, so a smaller cluster size must be used to find these lesions. This will reduce the overall point accuracy, but it is necessary in producing a good sensitivity. Very small lesions are difficult to locate with good accuracy, as they produce very small disease shadows that are often indistinguishable from the noisy error.

The symmetrical method sometimes produces a number of poor results. This is especially true with small lesions as they produce a limited size disease shadow. In these cases, a much smaller margin of error is needed to correctly detect the lesion. In Figure 4.9, the lesion and disease shadow is relatively small and difficult to distinguish from the error. Table 3.13 shows

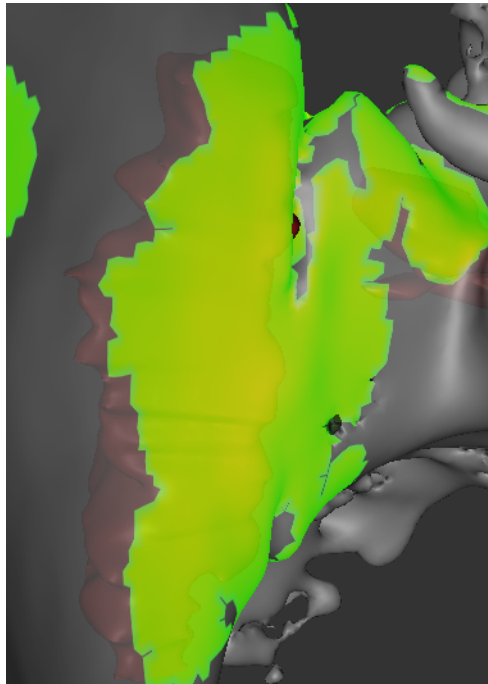


Fig. 4.8 A large 3D lesion (shown in red) produces a disease shadow (shown in green) with hundreds of points. This type of lesion is relatively easy to accurately locate as the cluster size can be set to a large number which removes much of the noisy error.

that the sensitivity significantly decreases in lesions that have a volume less than 0.5 cm^3 . Lesions that have a volume less than 0.1 cm^3 have a very low average sensitivity (0.48).

Disease shadows with greater differences in trabecular bone density are also easier to accurately detect as significantly diseased bony regions have a greater difference in bone density than the noisy error generally does. As shown in Figure 3.16, if a lesion can be detected using a greater HU cutoff value, it will improve the overall accuracy of the algorithm, hence as large as possible HU threshold should be used to detect the lesions. However, it is not always possible to detect lesions using a high HU so a lesser HU cutoff must be used at the cost of a reduced overall point accuracy. Some lesions only exhibit small changes in bone density, and in that case my method will not be able to detect these.

As the lesions have varying volumes and trabecular bone densities, certain sets of parameters work well on certain data sets and poorly on others. It is not possible to find a single set of parameters that produces the optimal point accuracy and sensitivity on all data sets as shown in Section 3.4.4. This requires checking a range of parameter values in order to optimally detect the lesions in each data set implying user interaction and some degree of interpretation.

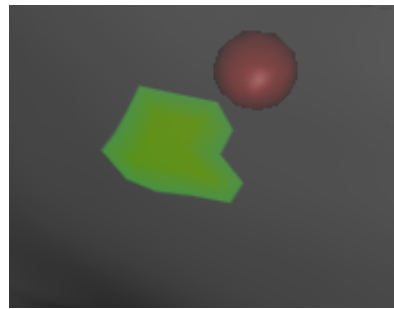


Fig. 4.9 The very small 3D lesion (shown in red) produces a disease shadow (shown in green) with a tiny number of points. A very small disease shadow such as this is very difficult to distinguish from error.

Using optimised parameters in the detection algorithms is ideal for producing the best results. Removing speckled error noise (achieved by setting a minimum cluster sizes or a minimum number of smoothing iterations), is important to validate the algorithm. Using as high as possible values of both cluster size and HU cut-off will produce better accuracy. However, each data set is significantly different and if either of these parameter values are set too high, there is a potential to miss legitimate (but either smaller or with lower change in density) lesions. Having an additional set of test data could confirm whether the optimised parameters would work well in general.

4.2.4 Discussion of overall results

As shown in Section 3.4.8, using a threshold cutoff of 125 HU, a minimum cluster size of 7 and 2 smoothing iterations produced an overall average point accuracy of 0.894 and an average sensitivity of 0.871 respectively when detecting lesions in a data set containing both sclerotic and lytic lesions using a five fold cross validation approach. The optimal parameters are the same for each fold which implies that using a testing set is less important. When detecting only sclerotic lesions, the assessment method produced an overall sensitivity of 0.906 (Section 3.4.10). It also produced an overall sensitivity of 0.852 when detecting only lytic regions. However, as described in Section 3.4.9, the labelling of the lesion type has not been verified by a radiologist. This is a shortcoming, although the marking was not possible to obtain due to constraints on the radiologist's time, especially during the pandemic.

Although the final classification of results in the various studies differ widely, the best state of the art machine learning technique [161], produced a sensitivity of 0.760 when detecting mixed sclerotic and lytic lesions. As mentioned in Section 1.1, mixed can refer to either a lesion type or a data set that contains both sclerotic and lytic lesions. Although my

data set does not have any lesions specifically labelled as mixed, my data set does include both sclerotic and lytic lesions. It achieved an overall sensitivity of 0.871 on these data sets.

When detecting only lytic lesions, the state of the art techniques produced a sensitivity of 0.88 [25] and 0.81 [161]. My method produced a sensitivity of 0.852 in detecting only lytic lesions. When detecting only sclerotic lesions, the best techniques produced a sensitivity of 0.97 [25], 0.81 [161] and 0.70 [105]. My assessment method achieved a sensitivity of 0.906 in detecting only sclerotic lesions. The size of the lesions used in all of these methods was similar, although none of the data sets used in any of the other studies were available to allow for a direct comparison.

However, these three state of the art methods are limited to finding lesions in the spinal column. The method described in this thesis finds lesions in the ribs, spinal column, pelvis and upper femurs. My technique also provides a 3D visualisation of the results that can be used to explain the results and to provide the locations of the lesions, and also provides an immediate way for a clinician to confirm in the CT data whether lesion detections are valid or not.

With regards to the number of false positives, all methods produce a significant number. It's difficult to compare between methods as my method uses a surface based approach while these other methods use a voxel based approach. However, Chmelik et al. [25] produced 3.4 to 7.8 false positives per vertebrae, while Yao et al. [161] produced a false positive rate of 0.81 (the false positive rate is defined as the number of false positives divided by the number of false positives added to the number of true negatives). Roth et al. [105] produced 3372 false positives and 935 true positives across 59 patients when detecting sclerotic lesions. My assessment method produced a false positive rate of 2.70% on a point labelling level. This is a greater false positive rate than existing methods, although this is a surfaced based measurement instead of a voxel based measurement.

My assessment method produces results comparable to the existing methods, although it is difficult to measure how well each of the results precisely compares to my results. It is also important to note that the sensitivities are based on detecting a small percentage of the actual diseased region (i.e. Chmelik et al. [25] required at least a 1% overlap between detected voxels and the lesion volume in order to mark a lesion as detected) and are not detecting the entirety of the diseased regions. This potentially poses a problem for determining the validity of the results available in the literature.

4.2.5 Discussion on volume calculation

As shown in Figure 3.6, the fit line for mapping disease shadow size to lesion volume follows the general trend of the data. However, a significant variance exists from individual

data points to the fit line. This is likely due to the varying shape of the bone constraining the lesion volume and disease shadows in different ways. As described in Section 3.2.3, calculating volume from disease shadow sizes produced an absolute percentage error of 52.2%. This shows that when calculating lesion volume using the disease shadow produced by the symmetrical assessment method, the volume will generally predict the correct value but the prediction could potentially have a significant variance from the true underlying lesion volume sizes.

This can be seen in Table 3.21, as the volume percentage difference ranges from 107.41% to 209.25%. It is also important to note that the data in this table is from the longitudinal experiments. The disease shadows used in calculating the lesion volume have been caused by the growth or shrinkage of the lesions.

4.2.6 Disease side selection

On a point basis, only 4.01% of all points that were incorrectly marked as diseased (as they were actually healthy) had their symmetrically mapped point labelled incorrectly as being healthy (as it was diseased). This demonstrates that the disease side selection algorithm does pick the correct side for the majority of the points.

As described in Section 3.4.9, a 25% overlap exists in ground truth symmetrical disease shadow points. Even though this is a relatively high percentage, the disease side selection method successfully picks the disease side correctly in the majority of the cases. This occurs because it is very unlikely that two disease shadows will have the same border even if they do overlap a significant amount. Also, although there is a 25% overlap, this may not always be between two of the same lesion type (i.e. lytic and lytic). If either lesion are of a different type, the disease side selection will be more accurate.

As the symmetrical assessment method provides a visualisation of potentially diseased regions, the sensitivity of this method could be immediately improved if the disease side selection was disregarded and both sides were offered to the user for inspection, comparison and disease selection. The cost of this would be the time required to manually inspect the extra regions, which may not be necessary as the disease side selection does a relatively good job at selecting the correctly diseased side. In any case, if one were to examine a selected side, the other side is highlighted at the same time, so it would be easy to manually check whether the side was not correct.

4.3 Discussion on general functionality

This algorithm is unique in its ability to transform noisy greyscale 2D images into a 3D model that can be used for displaying a potentially diseased region. It is the combination of the different individual elements that create the novel technique. Segmentation of bone is a reasonably common technique, although much less so when applied to a significant part of the whole body. However, taking the results from segmentation, overlaying the segmentation with trabecular bone density from CBM and symmetrically comparing the trabecular bone density is a completely novel combination of these components.

The symmetrical technique is useful in finding differences in trabecular bone density between symmetric parts of the bone structure. The applications using this method are potentially extensive, as anything that produces abnormalities in trabecular bone densities could potentially be found. This could include finding any other abnormalities, other than those purely limited to cancer. Broken or injured bones, for example, can potentially be found using this method. However, symmetric differences in trabecular bone density are especially prevalent in metastatic bone disease, which has been the focus of this project, as the secondary spread often causes severe changes in trabecular bone densities.

4.3.1 Usefulness pertaining to metastatic bone disease

The symmetrical method described in this thesis is capable of detecting MBD numerically in the form of a predicted lesion volume size and as the number of points in a disease shadow. This numerical output is both novel and an important distinction, as the current method of diagnosing MBD is visual. The current diagnostic technique, as described in Section 1.2, requires extensive training of a medical professional and identification skills of matching a current radiographic image to what would be regarded as physiological rather than pathological variation. While this is the gold standard of disease diagnosis, my algorithm provides a secondary method of identifying the disease by comparing trabecular bone density directly, which reduces the reliance on human methodology and also reduces error and oversight. My method offers a simple and numerical way to evaluate indicators of disease with regard to differences in trabecular bone density.

When a radiologist looks at an image to diagnose disease, they are taking the knowledge they previously learned in their medical training and applying it to the radiographic image at hand. While this usually provides an effective diagnosis (depending on the skill and experience of the radiologist), there is always an element of human error or oversight.

My approach with this algorithm is to use the numerical data itself to determine the presence and extent of the disease. The number of points in each disease shadow can be

totalled to determine the approximate surface area of each shadow. This area measurement can be used to find an approximate lesion volume based on the 2 to 3 relationship between surface area and volume. This is not a precise number. However, when combined with the total number of disease shadows in a data set, these values provide a secondary numerical metric that could be useful to a radiologist, especially if the radiologist overlooked a lesion. This algorithm could potentially be used to provide both a numerical description of the extent of the disease as well as an automated secondary check, in order to remove some of the dependency on potential human error. This method is not intended to replace a radiologist, but rather to support them by providing them with another tool where the expert knowledge of medical education and experience is combined with scientific and mathematical calculations. This, in turn, creates the best possible diagnostic situation.

4.3.2 Longitudinal analysis – changes in trabecular bone density over time

According to Dr. Barrett, accurately tracking disease progression in MBD is usually a far more onerous problem for a radiologist to overcome than the initial diagnosis is. As explained in Section 1.5, this process consists of imaging a patient at multiple time-points. The data sets are then viewed side by side, where the radiologist evaluates them using a visual comparison, looking for small changes between the scans.

Clinicians broadly categorise tumour growth into four states using the response evaluation criteria in solid tumours (RECIST) rules. RECIST refers to a set of published rules used to assess tumour burden in order to provide an objective assessment of the response to therapy [88]. These states include a complete response (the disappearance of all lesions), a partial response (less than 30% decrease in sum of all target lesions), a stable disease and a progressive disease.

Potentially, this is where my algorithm could provide the most benefit for the medical community. Using the registration techniques described in Section 2.6, my algorithm can map the 3D bone segmentations taken from the scans to each other and directly compare the difference in trabecular bone densities, providing a much more automatic and possibly more precise visualisation of results for a radiologist to use when comparing the scans. Any combination of possible baseline and follow-up scans can alternatively be compared with each other, further increasing the usefulness of the longitudinal analysis technique.

This technique can also provide numerical values (of differences in disease shadow size as well as changes in volume) that represents the amount of growth or stagnancy with regard to the progression of the disease at different time-points. This is a novel and important metric

as radiologists are unable to quantify a precise numerical value for the baseline and follow-up differences in diseased scans without painstakingly measuring areas in the CT scan by hand. They usually evaluate the differences and mark them with varying degrees of progression, but are unable to easily and precisely measure the sum total of differences.

While similar approaches [94, 108, 141] have been used before in order to compare Hounsfield units at the bone's surface, comparing the trabecular bone density measurements in 3D has not been applied to this longitudinal technique before. Most of the MBD is located within the trabecular bone, thus using the trabecular bone density measurements is an essential component to making this longitudinal analysis work correctly—which is a more effective method of evaluating the spread of the disease.

4.4 Visualisations

While my algorithm can produce numerical reports of the bone in a data set being either diseased or healthy, these numerical values can be projected onto the 3D bony segmented surfaces, creating novel visualisations of the disease. The CT data sets contain noisy, greyscale 2D images, which a radiologist is required to examine slice by slice. This can often consist of inspecting hundreds of images, looking for minuscule changes. By encapsulating the bone into a 3D mesh, the bony data set can be viewed in its entirety, and the disease can be projected onto this 3D bone model.

This surface visualisation provides a unique way of viewing the skeleton, both in terms of locating the disease, as well as viewing the size and spread. In this way, the 3D visualisation has the potential to be an incredibly beneficial tool, both for the radiologist, as it provides them a 3D imaging technique that helps to summarise the data, but also for the patients who will most likely find looking at a 3D model of the disease easier to understand than looking at noisy greyscale CT images as the diseased and healthy regions can be coloured with any user-defined colour scheme.

Although this surface visualisation has not been tested in a clinical setting, an important next step in this research would be to use this method in a clinical setting to demonstrate its benefits.

The direct trabecular bone density and symmetrical differences in trabecular bone density can also be mapped to colour values and displayed on the 3D mesh, providing more novel methods of visualisation that could be useful in evaluating disease.

When using a 3D model of the bone, the underlying CT data is not lost but incorporated into the visualisations, as the 3D skeletal mesh is mapped to the CT data and can be displayed with a user click. This process is described in Section 2.5.4. Both the original CT slices can

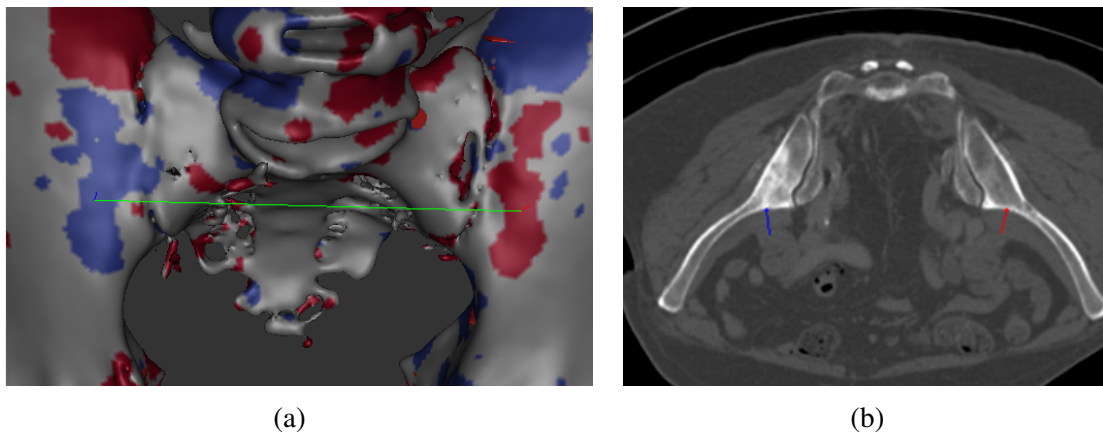


Fig. 4.10 An example of mapping the 3D visualisations back to the original CT images. On the left, a user has clicked on a diseased region (shown by the red surface normal line). On the right, this diseased data is shown in the resliced CT data (which is the plane defined by the vector found by averaging the red and blue surface normal vectors and by the vector of the green line connecting the two points). The CT data is from the Addenbrooke's Hospital CT Data Collection. Image is a screenshot taken of the custom image viewer.

be displayed as well as a resliced plane defined by the bony symmetry. This can be seen in Figure 4.10 where the diseased region has been projected onto the 3D surface (shown on the left) and also mapped back to the original CT data (resliced and shown on the right).

4.4.1 Use in a clinical setting

In a clinical setting, my algorithm has the potential to provide significant benefits. As scanning patients in a CT machine is a very common medical imaging technique used for diagnosing a wide range of diseases and injuries, this algorithm could be run on every CT scan produced in a clinic in order to check for metastatic bone disease with little additional clinician time required. Again, this assessment method requires no manual interaction at all in the entire assessment process.

The output of each run could produce a report containing the results of whether MBD was likely to be present in the scan. Numerical volume results could indicate the extent of the disease describing the size and possible number of lesions. As my program can run on a normal desktop computer (with an acceptable graphics card), this executable could easily run in the background on a laboratory machine. It could notify the clinic if any disease was found and continue silently otherwise. If disease was found, clinicians could interpret the report and could notify an appropriate radiologist to take a closer look at the CT data. Using this method, disease has the potential to be diagnosed earlier than it would normally have been, potentially leading to a better outcome for the patient.

Furthermore, producing numerical data describing the disease both in assessing the extent of MBD in a patient as well as the disease progression is valuable in a clinical setting. This allows for more precise labelling of the disease and a more accurate comparison between patients and cases. The most usual numerical labelling would most likely be when assessing disease across time-points, as this is more difficult for a radiologist to do.

Additionally, this method could potentially lead to a number of improvements in a clinical setting that include a reduction in the amount of time it takes for a radiologist to assess MBD. It can also lead to an improved confidence in the assessment and a more thorough screening of the disease (by running it on every CT scan).

4.4.2 Clinical pilot study

For an initial pilot study, a single case was selected at random. It was loaded into the image viewer and nine screen shots were taken of different views. These images showed the 3D bone segmentation overlaid with trabecular bone density colour mapping as well as the 3D bone segmentation overlaid with the symmetrical differences in trabecular bone density between the left and right sides. Screenshots were also taken of potentially diseased areas (areas of strong symmetrical difference in trabecular bone density) along with the original and resliced CT images (from a user click) that intersected these potentially diseased regions.

These screenshot images were presented to Dr. Barrett for feedback on the potential usefulness of the visualisations. Dr. Barrett noted that the 3D segmentations overlaid with trabecular bone density colour mapping as well as 3D segmentations overlaid with the trabecular bone density symmetrical differences were particularly useful in diagnostic analysis. He also commented on the resliced (axial) CT images (intersecting through the diseased regions) and said that creating a resliced image from a user click on a symmetrical region could be very beneficial. However he wanted the ability to scroll through the axial images based on the initial resliced image's orientation—something that can be easily added.

This small study shows that the viewer developed for this thesis would be useful in a clinical setting. However additional work would be needed to expand the functionality of the reslicing visualisation feature to make it more beneficial for a radiologist's ease of use. Based on this small study, a larger clinical study including many more cases would be helpful in order to gather additional data points on the potential usefulness of this tool.

4.4.3 Limitations

There are limitations when using the symmetrical method to detect lesions. Most obviously, if the lesions form symmetrically, the hardened bone density in each symmetrical area will

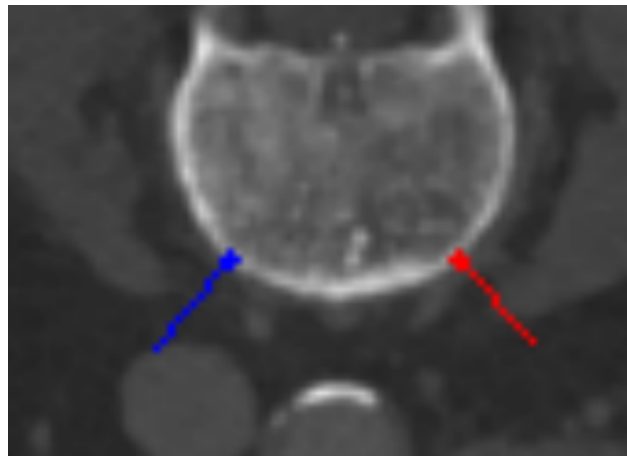


Fig. 4.11 The entire vertebrae shown in this image is diseased. The hardened trabecular bone density on each side cancels the other out so the region is not marked as diseased. The underlying CT data is from the Addenbrooke's Hospital CT Data Collection.

cancel out each other and the region will not be marked as a diseased region. While this is uncommon, it does happen, most often in the vertebrae, as seen in Figure 4.11. In this image, the entire vertebra has been diseased.

Also, human bones do not develop in an exactly symmetrical way. In Figure 4.12, it is obvious that the bone is not completely symmetrical in this case, which can introduce error into the registration results. This could either be the result of physiological difference between left and right sides, or the result of pathology, or from poor segmentation.

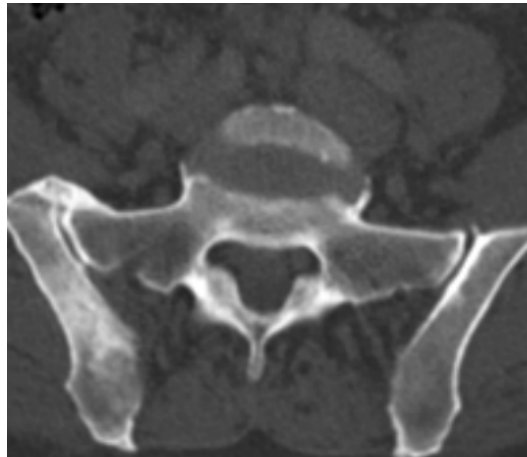


Fig. 4.12 The human skeleton does not develop in exactly a symmetric way. The bone within this CT slice is not perfectly symmetric, and this lack of symmetry is likely due to development as well as to disease. Additionally, the slice angle could be another cause of asymmetry with CT images. The CT data is from the Addenbrooke's Hospital CT Data Collection.

Chapter 5

Conclusions

5.1 Introduction

The purpose of this project was to investigate using a symmetry-based method to compare trabecular bone density in order to locate metastatic bone disease and to track its spread. A secondary component was to compare the trabecular bone density of CT scans taken at different time-points in order to track the spread of MBD.

A significant number of components exist in this project including a fully automatic segmentation technique, a deformable registration technique, the creation of a completely symmetric atlas, parameter optimisations and a 3D viewer capable of comparing trabecular bone density and overlaying the diseased regions upon a 3D segmentation mesh.

5.2 Discussion and conclusions

The approach used in this thesis to locate metastatic bone disease is completely novel as it varies greatly from the machine learning approaches described in Chapter 1. It offers benefits over these other methods as it provides a 3D visualisation of the diseased regions and disease spread that can be used to verify the results whilst minimising the need for manual input.

While it is difficult to directly compare results to existing methods as they solve a slightly different problem, the methods described in this thesis improve upon these existing methods by providing novel visualisations of the disease while maintaining comparable sensitivities. This method can find lesions in the ribs, spine, pelvis and upper femurs and is not limited to the spinal column, as the other methods are.

The symmetrical assessment method can accurately find lesions greater than 0.1 cm^3 in the spine, pelvis and upper femurs. However, the accuracy of the ribs is much poorer due to

their shape, size and location. Also, widespread disease in the vertebrae is difficult for the symmetrical assessment method to find.

Improvements can be made to this algorithm to improve its accuracy and sensitivity and to decrease the number of false positives. However, it is a fully working algorithm that provides a significant milestone in lesion detection in MBD and can be used as a benchmark for others to compare results against.

5.3 Improvements

There are a number of improvements that could be made to the different subsections of the algorithm that might improve the accuracy of each section as well as the systemic accuracy as a whole. These improvements cover issues with the symmetry, the registration, limitations to Cortical Bone Mapping (CBM), in addition to the need to speed up the algorithm itself. These could serve for future research projects that would hopefully add to the ability to accurately diagnose bone disease.

5.3.1 Improvements to segmentation

Segmentation of bone and CT, as described in Section 2.2.1, is a problem that has not yet been fully solved. Although the results for the algorithm's segmentation are generally good, it can fail in many places, especially in cases where the bone is not distinct, and has Hounsfield units that are similar to surrounding tissue or contains irregularities such as the CT table, metal implants or contrast agents. A significant amount of additional research would be necessary in this area in order to produce a perfect, automatic, and fast segmentation of the bone.

In order to improve the segmentation, it is important to address the partial volume effect. This would be reduced by increasing the CT resolution, which would also enable better segmentation accuracies. It may be possible to combine the method of hysteresis thresholding, with a state-of-the-art neural network technique (such as Klein et al. [71]) to help improve the deformable registration accuracy. If the segmentation is improved, it will impact all future downstream algorithmic use in this program, which should improve the overall accuracy and sensitivity of the algorithm. Producing a very accurate segmentation will also improve the CBM mapping, which will improve the trabecular bone density measurements—this is an absolutely vital step to improving overall accuracy and sensitivity.

Having a more precise segmentation will reduce the noise error in the visualisation, which will allow the cluster size parameter to be set lower and the threshold cut-off parameter to

also be set lower. Both of these will improve the sensitivity without sacrificing accuracy, and without increasing the number of false positives.

It would also be helpful if more CT data were available, especially if the bone within the CT data was manually segmented with greater accuracy. This would produce a more robust ground truth and would allow for more accurate experiments and produce more data (for instance) for machine learning algorithms on which to train. Additionally, a more robust test could be achieved if each lesion was marked with a type (i.e. sclerotic or lytic).

5.3.2 Improvements to registration

Remaining errors in registration are from a number of sources, including severe differences between the atlas and the segmentation – which can include either strong differences in anatomy, or different sizes and scanned areas. Sometimes there are difficulties in initial placement that can produce poor results, since the registration error will not be correctly minimised from the wrong starting point.

There are a number of things that might be done to improve registration, including using multiple atlases, where the atlas that best registers to the segmentation is used and the others discarded. Whitmarsh et al. [152] used a similar multi-atlas segmentation technique. Also, the atlas vertebrae registration might be improved using a flexible hinge-type model where the vertebrae are free to move around with some degree of freedom within a joint angle setting. This could provide a more structured registration for the vertebrae, as they sometimes float and match to an incorrect vertebra. Equally, the additional restriction of the hinges could prevent a bone registering correctly and actually make the overall registration less precise.

The part of registration that is by far the most problematic is the ribs. This is because the ribs have large variations in their shape, size, number, and general anatomy when comparing the segmentation to the atlas. Although this issue matters less in this project, as almost all the diseased bone is located closer to the prostate, lesions in this area are much more common in other types of cancer, such as breast cancer where the ribs are more closely located to the primary site of the disease.

5.3.3 Improvements to CBM

An additional area of improvement that needs to be addressed is the CBM measurements. CBM can be improved in handling thin regions of the bone, which occurs in various parts of the pelvis, as well as in the ribs. As CBM expects the bone to be a certain length, the modelling process is not accurate in very thin areas. Because of this, my algorithm often labels these thin areas as diseased when they are not. To compensate for this, the algorithm

checks for thin areas and ignores these regions. However, a better solution would be to adjust the CBM modelling process in some way when a thin area is found.

Another possible way to improve the very thin regions would be to use the CBM line measurements directly. Line data was used in the experiments in Section 3.4.11 to see if this would achieve better results than CBM. Using line data bypasses the CBM modelling approach but has shown to work significantly less well than using CBM. However, it may be possible to achieve better results if this approach is only used when the bone is very thin, and hence when the CBM modelling process does not produce valid trabecular density measurements.

Significant improvement is needed in detecting small lesions, especially those with low differences in trabecular bone density. As mentioned in Section 3.4.5, speckled error noise within the program is often indistinguishable from the signal produced by legitimately diseased regions. In order to detect these, the overall accuracy of all the subsections of the algorithm would need to be improved.

5.4 Future Extensions

As this is a novel technique, there are a large number of related future extensions that can be explored. This section provides an overview of a number of possible areas that could be extensions of the work described in this thesis.

This algorithm is capable of finding any changes within the skeletal system, and so may be applied to any applications where there are changes in bone density. The obvious next application is detecting MBD from other types of cancer such as breast cancer. Slight modifications would need to be made, as the lesions in a breast cancer circumstance will more likely be focused around the chest, and will have a more common prevalence in the ribs and upper vertebra.

There are more sclerotic lesions than lytic lesions in my data set. Since my algorithm is better at finding sclerotic lesions, a new way of detecting lytic lesions could be incorporated into the algorithm. Also, a new atlas will be needed for breast cancer, as the current atlas is of a male skeleton. In the case of breast cancer, the patients are nearly all women, thus a female skeleton atlas will be essential. Also, the atlas would need to be centred around the chest instead of the prostate as a close proximity to the primary site of cancer is needed.

Another area of investigation would be to look into using this technique to detect and track osteoporosis. Osteoporosis weakens the bone structure over time, and so as long as the osteoporosis does not weaken the bone in a completely symmetric and uniform manner, which is unlikely, this algorithm should be able to determine the weakened areas of the bone,

creating an opportunity for diagnosis. This process could also be used on longitudinal CT scans in order to find and track the disease progression of osteoporosis. Again, the atlas will have to be adjusted depending on the sex of the patient, and the location of the bony regions.

An additional extension of this project would be to show how symmetric a general human skeleton actually is, as a plane of symmetry can be defined, optimised and used to create a perfectly symmetric segmentation by projecting its points across the plane of symmetry. By comparing the points in this projection to the original segmentation, the amount of difference between the closest points will provide a measure of the amount of symmetry (or the lack thereof) in the skeleton.

References

- [1] Allen, B., Curless, B., and Popović, Z. (2002). Articulated body deformation from range scan data. *ACM Transactions on Graphics (TOG)*, 21(3):612–619.
- [2] Allen, B., Curless, B., and Popović, Z. (2003). The space of human body shapes: reconstruction and parameterization from range scans. *ACM transactions on graphics (TOG)*, 22(3):587–594.
- [3] Alperin, R. C. (1987). Heron’s area formula. *The College Mathematics Journal*, 18(2):137–138.
- [4] Amberg, B., Romdhani, S., and Vetter, T. (2007). Optimal step nonrigid ICP algorithms for surface registration. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE.
- [5] Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., and Davis, J. (2005). Scape: shape completion and animation of people. In *ACM SIGGRAPH 2005 Papers*, pages 408–416. ACM.
- [6] Arabi, H. and Zaidi, H. (2017). Comparison of atlas-based techniques for whole-body bone segmentation. *Medical image analysis*, 36:98–112.
- [7] Bae, H.-J., Hyun, H., Byeon, Y., Shin, K., Cho, Y., Song, Y. J., Yi, S., Kuh, S.-U., Yeom, J. S., and Kim, N. (2020). Fully automated 3D segmentation and separation of multiple cervical vertebrae in CT images using a 2D convolutional neural network. *Computer Methods and Programs in Biomedicine*, 184:105119.
- [8] Baiker, M., Milles, J., Dijkstra, J., Henning, T. D., Weber, A. W., Que, I., Kaijzel, E. L., Löwik, C. W., Reiber, J. H., and Lelieveldt, B. P. (2010). Atlas-based whole-body segmentation of mice from low-contrast micro-ct data. *Medical image analysis*, 14(6):723–737.
- [9] Baiker, M., Snoeks, T. J., Kaijzel, E. L., Que, I., Dijkstra, J., Lelieveldt, B. P., and Löwik, C. W. (2012). Automated bone volume and thickness measurements in small animal whole-body microct data. *Molecular imaging and biology*, 14(4):420–430.
- [10] Baiker, M., Staring, M., Löwik, C. W., Reiber, J. H., and Lelieveldt, B. P. (2011). Automated registration of whole-body follow-up microct data of mice. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 516–523. Springer.

- [11] Bartalucci, C., Furferi, R., Governi, L., and Volpe, Y. (2018). A survey of methods for symmetry detection on 3D high point density models in biomedicine. *Symmetry*, 10(7):263.
- [12] Baumgart, B. G. (1975). A polyhedron representation for computer vision. In *Proceedings of the May 19-22, 1975, national computer conference and exposition*, pages 589–596.
- [13] Besl, P. J. and McKay, N. D. (1992). Method for registration of 3-D shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. International Society for Optics and Photonics.
- [14] Bieth, M., Peter, L., Nekolla, S. G., Eiber, M., Langs, G., Schwaiger, M., and Menze, B. (2017). Segmentation of skeleton and organs in whole-body CT images via iterative trilateration. *IEEE transactions on medical imaging*, 36(11):2276–2286.
- [15] Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585.
- [16] Bookstein, F. L. (1997). *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press.
- [17] Brahim, W., Mestiri, M., Betrouni, N., and Hamrouni, K. (2016). Semi-automated rib cage segmentation in CT images for mesothelioma detection. In *2016 International Image Processing, Applications and Systems (IPAS)*, pages 1–6. IEEE.
- [18] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6):394–424.
- [19] Bronner, F. and Farach-Carson, M. C., editors (2009). *Bone and cancer*. Topics in bone biology ; v. 5. Springer, London.
- [20] Brown, B. J. and Rusinkiewicz, S. (2007). Global non-rigid alignment of 3-D scans. In *ACM SIGGRAPH 2007 papers*, pages 21–es. Association for Computing Machinery (ACM).
- [21] Buie, H. R., Campbell, G. M., Klinck, R. J., MacNeil, J. A., and Boyd, S. K. (2007). Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for in vivo Micro-CT bone analysis. *Bone*, 41(4):505–515.
- [22] Burns, J. E., Yao, J., Wiese, T. S., Muñoz, H. E., Jones, E. C., and Summers, R. M. (2013). Automated detection of sclerotic metastases in the thoracolumbar spine at CT. *Radiology*, 268(1):69–78.
- [23] Buzug, T. M. (2008). *Computed tomography : from photon statistics to modern cone-beam CT* / Thorsten M. Buzug. Springer, Berlin.
- [24] Chen, H., Shen, C., Qin, J., Ni, D., Shi, L., Cheng, J. C., and Heng, P.-A. (2015). Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks. In *International conference on medical image computing and computer-assisted intervention*, pages 515–522. Springer.

- [25] Chmelik, J., Jakubicek, R., Walek, P., Jan, J., Ourednicek, P., Lambert, L., Amadori, E., and Gavelli, G. (2018). Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data. *Medical image analysis*, 49:76–88.
- [26] Cho, J., Lee, K., Shin, E., Choy, G., and Do, S. (2015). How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? *arXiv preprint arXiv:1511.06348*.
- [27] Chu, C., Chen, C., Liu, L., and Zheng, G. (2015). Facts: fully automatic CT segmentation of a hip joint. *Annals of biomedical engineering*, 43(5):1247–1259.
- [28] Chu, G. H., Lo, P., Kim, H. J., Lu, P., Ramakrishna, B., Gjertson, D., Poon, C., Auerbach, M., Goldin, J., and Brown, M. S. (2012). Automated segmentation of tumors on bone scans using anatomy-specific thresholding. In *Medical Imaging 2012: Computer-Aided Diagnosis*, volume 8315, page 83150F. International Society for Optics and Photonics.
- [29] Chui, H. and Rangarajan, A. (2003). A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2-3):114–141.
- [30] Coleman, R. and Rubens, R., editors (1992). *Metastatic Bone Disease*. Parthenon, Carnforth.
- [31] Cook, G. J. (2010). PET and PET/CT imaging of skeletal metastases. *Cancer Imaging*, 10(1):153.
- [32] Department of Surgery, University of Cambridge (2021). Cambridge urology translational research and clinical trials office (CUTRACT). <https://surgery.medschl.cam.ac.uk/divisions-and-groups/academic-urology-group/cambridge-urology-translational-research-and-clinical-trials-office-cutract/>.
- [33] Desbrun, M., Meyer, M., Schröder, P., and Barr, A. H. (1999). Implicit fairing of irregular meshes using diffusion and curvature flow. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 317–324.
- [34] Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.
- [35] Diel, I. J., Kaufman, M., and Bastert, G. (1994). *Metastatic Bone Disease: Fundamental and Clinical Aspects*. Springer-Verlag.
- [36] Doi, A. and Koide, A. (1991). An efficient method of triangulating equi-valued surfaces by using tetrahedral cells. *IEICE TRANSACTIONS on Information and Systems*, 74(1):214–224.
- [37] Drake, R. L. et al. (2014). *Gray’s anatomy for students*. Churchill Livingstone, third edition.
- [38] Dutta, S., Das, B., and Kaushik, S. (2019). Assessment of optimal deep learning configuration for vertebrae segmentation from CT images. In *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications*, volume 10954, page 109541A. International Society for Optics and Photonics.

- [39] Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
- [40] Elmoutaouakkil, A., Peyrin, F., Elkafi, J., and Laval-Jeantet, A.-M. (2002). Segmentation of cancellous bone from high-resolution computed tomography images: influence on trabecular bone measurements. *IEEE transactions on medical imaging*, 21(4):354–362.
- [41] Faugeras, O. D. and Hebert, M. (1986). The representation, recognition, and locating of 3-D objects. *The international journal of robotics research*, 5(3):27–52.
- [42] Feldmar, J. and Ayache, N. (1996). Rigid, affine and locally affine registration of free-form surfaces. *International journal of computer vision*, 18(2):99–119.
- [43] Fu, Y., Liu, S., Li, H. H., and Yang, D. (2017). Automatic and hierarchical segmentation of the human skeleton in CT images. *Physics in Medicine & Biology*, 62(7):2812.
- [44] Furnstahl, P., Fuchs, T., Schweizer, A., Nagy, L., Székely, G., and Harders, M. (2008). Automatic and robust forearm segmentation using graph cuts. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 77–80. IEEE.
- [45] Gassman, E. E., Powell, S. M., Kallemeyn, N. A., DeVries, N. A., Shivanna, K. H., Magnotta, V. A., Ramme, A. J., Adams, B. D., and Grosland, N. M. (2008). Automated bony region identification using artificial neural networks: reliability and validation measurements. *Skeletal radiology*, 37(4):313–319.
- [46] Gelaude, F., Vander Sloten, J., and Lauwers, B. (2006). Semi-automated segmentation and visualisation of outer bone cortex from medical images. *Computer methods in biomechanics and biomedical engineering*, 9(1):65–77.
- [47] Ginde, A. A., Foianini, A., Renner, D. M., Valley, M., and Camargo, Jr, C. A. (2008). Availability and quality of computed tomography and magnetic resonance imaging equipment in US emergency departments. *Academic emergency medicine*, 15(8):780–783.
- [48] Glocker, B., Feulner, J., Criminisi, A., Haynor, D. R., and Konukoglu, E. (2012). Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 590–598. Springer.
- [49] Glocker, B., Zikic, D., Konukoglu, E., Haynor, D. R., and Criminisi, A. (2013). Vertebrae localization in pathological spine CT via dense classification from sparse annotations. In *International conference on medical image computing and computer-assisted intervention*, pages 262–270. Springer.
- [50] Gunz, P. and Mitteroecker, P. (2013). Semilandmarks: a method for quantifying curves and surfaces. *Hystrix, the Italian journal of mammalogy*, 24(1):103–109.
- [51] Gunz, P., Mitteroecker, P., and Bookstein, F. L. (2005). Semilandmarks in three dimensions. In *Modern morphometrics in physical anthropology*, pages 73–98. Springer.
- [52] Hamaoka, T., Madewell, J. E., Podoloff, D. A., Hortobagyi, G. N., and Ueno, N. T. (2004). Bone imaging in metastatic breast cancer. *Journal of Clinical Oncology*, 22(14):2942–2953.

- [53] Hammon, M., Dankerl, P., Tsymbal, A., Wels, M., Kelm, M., May, M., Suehling, M., Uder, M., and Cavallaro, A. (2013). Automatic detection of lytic and blastic thoracolumbar spine metastases on computed tomography. *European radiology*, 23(7):1862–1870.
- [54] Hangartner, T. N. (2007). Thresholding technique for accurate analysis of density and geometry in QCT, pQCT and muCT images. *Journal of Musculoskeletal and Neuronal Interactions*, 7(1):9.
- [55] Hardisty, M., Gordon, L., Agarwal, P., Skrinskas, T., and Whyne, C. (2007). Quantitative characterization of metastatic disease in the spine. Part I. semiautomated segmentation using atlas-based deformable registration and the level set method. *Medical physics*, 34(8):3127–3134.
- [56] Hemke, R., Buckless, C. G., Tsao, A., Wang, B., and Torriani, M. (2020). Deep learning for automated segmentation of pelvic muscles, fat, and bone from CT studies for body composition assessment. *Skeletal Radiology*, 49(3):387–395.
- [57] Holzinger, A. (2018). From machine learning to explainable AI. In *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, pages 55–66. IEEE.
- [58] Holzinger, A., Kieseberg, P., Weippl, E., and Tjoa, A. M. (2018). Current advances, trends and challenges of machine learning and knowledge extraction: from machine learning to explainable AI. In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, pages 1–8. Springer.
- [59] Huang, Q.-X., Adams, B., Wicke, M., and Guibas, L. J. (2008). Non-rigid registration under isometric deformations. In *Computer Graphics Forum*, volume 27, pages 1449–1457. Wiley Online Library.
- [60] Huang, S.-F. and Chiang, K.-H. (2012). Automatic detection of bone metastasis in vertebrae by using CT images. In *Proceedings of the World Congress on Engineering*, volume 2, pages 1166–1171.
- [61] Jan, J., Novosadova, M., Demel, J., Ouředníček, P., Chmelík, J., and Jakubíček, R. (2015). Combined bone lesion analysis in 3D CT data of vertebrae. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6374–6377. IEEE.
- [62] Janssens, R., Zeng, G., and Zheng, G. (2018). Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3D fully convolutional networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 893–897. IEEE.
- [63] Jasmin, C. et al., editors (2005). *Textbook of bone metastases*. Wiley, Chichester.
- [64] Jerebko, A. K., Schmidt, G., Zhou, X., Bi, J., Anand, V., Liu, J., Schoenberg, S., Schmuecking, I., Kiefer, B., and Krishnan, A. (2007). Robust parametric modeling approach based on domain knowledge for computer aided detection of vertebrae column metastases in MRI. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 713–724. Springer.

- [65] Kang, Y., Engelke, K., and Kalender, W. A. (2003). A new accurate and precise 3-D segmentation method for skeletal structures in volumetric CT data. *IEEE transactions on medical imaging*, 22(5):586–598.
- [66] Kelm, B. M., Wels, M., Zhou, S. K., Seifert, S., Suehling, M., Zheng, Y., and Comaniciu, D. (2013). Spine detection in CT and MR using iterated marginal space learning. *Medical image analysis*, 17(8):1283–1292.
- [67] Khmelinskii, A., Baiker, M., Chen, X. J., Reiber, J. H., Henkelman, R. M., and Lelieveldt, B. P. (2010). Atlas-based organ & bone approximation for ex-vivo μ MRI mouse data: a pilot study. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1197–1200. IEEE.
- [68] Khmelinskii, A., Baiker, M., Kaijzel, E. L., Chen, J., Reiber, J. H., and Lelieveldt, B. P. (2011a). Articulated whole-body atlases for small animal image analysis: construction and applications. *Molecular imaging and biology*, 13(5):898–910.
- [69] Khmelinskii, A., Baiker, M., Kok, P., de Swart, J., Reiber, J. H., de Jong, M., and Lelieveldt, B. P. (2011b). Atlas-based articulated skeleton segmentation of μ spect mouse data. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 437–440. IEEE.
- [70] Khmelinskii, A., Groen, H. C., Baiker, M., de Jong, M., and Lelieveldt, B. P. (2012). Segmentation and visual analysis of whole-body mouse skeleton microSPECT. *PloS one*, 7(11).
- [71] Klein, A., Warszawski, J., Hillengaß, J., and Maier-Hein, K. H. (2019). Automatic bone segmentation in whole-body CT images. *International journal of computer assisted radiology and surgery*, 14(1):21–29.
- [72] Klinder, T., Ostermann, J., Ehm, M., Franz, A., Kneser, R., and Lorenz, C. (2009). Automated model-based vertebra detection, identification, and segmentation in CT images. *Medical image analysis*, 13(3):471–482.
- [73] Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada.
- [74] Kok, P., Baiker, M., Hendriks, E. A., Post, F. H., Dijkstra, J., Lowik, C. W., Lelieveldt, B. P., and Botha, C. P. (2010). Articulated planar reformation for change visualization in small animal imaging. *IEEE transactions on visualization and computer graphics*, 16(6):1396–1404.
- [75] Krčah, M., Székely, G., and Blanc, R. (2011). Fully automatic and fast segmentation of the femur bone from 3D-CT images with no shape prior. In *2011 IEEE international symposium on biomedical imaging: from nano to macro*, pages 2087–2090. IEEE.
- [76] Lay, N., Liu, D., Nogues, I., Summers, R. M., et al. (2016). Accurate 3D bone segmentation in challenging CT images: Bottom-up parsing and contextualized optimization. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10. IEEE.

- [77] Lee, M. J., Hong, H., Shim, K. W., and Park, S. (2019a). Orbital bone segmentation in head and neck CT images using multi-gray level fully convolutional networks. In *Medical Imaging 2019: Image Processing*, volume 10949, page 109493D. International Society for Optics and Photonics.
- [78] Lee, S., Lee, M. J., Hong, H., Shim, K. W., and Park, S. (2019b). Automatic segmentation of the orbital bone in 3D maxillofacial CT images with double-bone-segmentation network. In *International Forum on Medical Imaging in Asia 2019*, volume 11050, page 110500N. International Society for Optics and Photonics.
- [79] Lessmann, N., van Ginneken, B., and Išgum, I. (2018). Iterative convolutional neural networks for automatic vertebra identification and segmentation in CT images. In *Medical Imaging 2018: Image Processing*, volume 10574, page 1057408. International Society for Optics and Photonics.
- [80] Levoy, M. (1988). Display of surfaces from volume data. *IEEE Computer graphics and Applications*, 8(3):29–37.
- [81] Li, H., Adams, B., Guibas, L. J., and Pauly, M. (2009). Robust single-view geometry and motion reconstruction. *ACM Transactions on Graphics (ToG)*, 28(5):1–10.
- [82] Li, X., Gong, Z., Yin, H., Zhang, H., Wang, Z., and Zhuo, L. (2020). A 3D deep supervised densely network for small organs of human temporal bone segmentation in CT images. *Neural Networks*.
- [83] Liao, M., Zhang, Q., Wang, H., Yang, R., and Gong, M. (2009). Modeling deformable objects from a single depth camera. In *2009 IEEE 12th International Conference on Computer Vision*, pages 167–174. IEEE.
- [84] Liu, S., Xie, Y., and Reeves, A. P. (2017a). Individual bone structure segmentation and labeling from low-dose chest CT. In *Medical Imaging 2017: Computer-Aided Diagnosis*, volume 10134, page 1013444. International Society for Optics and Photonics.
- [85] Liu, X., Wu, Y., and Wang, B. (2017b). Spinal CT image segmentation based on level set method. In *2017 36th Chinese Control Conference (CCC)*, pages 10956–10961. IEEE.
- [86] Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *ACM siggraph computer graphics*, 21(4):163–169.
- [87] Maya (2021). Maya, 3D computer animation, modeling, simulation, and rendering software. <https://www.autodesk.com/products/maya/overview?term=1-YEAR&tab=subscription>.
- [88] McHugh, K. and Kao, S. (2003). Response evaluation criteria in solid tumours (RECIST): problems and need for modifications in paediatric oncology? *The british journal of radiology*, 76(907):433–436.
- [89] McLeod, R. and Stephens, D. (1979). Computed tomography of pelvic musculoskeletal neoplasm. *Contemp Orthop*, 1:36–41.
- [90] Moré, J. J. (1978). The Levenberg-Marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer.

- [91] Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- [92] Myronenko, A. and Song, X. (2010). Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2262–2275.
- [93] Nguyen, C., Havlicek, J., Duong, Q., Vesely, S., Gress, R., Lindenberg, L., Choyke, P., Chakrabarty, J. H., and Williams, K. (2016). An automatic 3D CT/PET segmentation framework for bone marrow proliferation assessment. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 4126–4130. IEEE.
- [94] Oh, J., Kim, G., Lee, J., Cheon, M., Park, Y., Kim, S., Yi, J., and Lee, H. Y. (2017). Automated detection of bone metastatic changes using serial CT scans. *Computerized Medical Imaging and Graphics*, 58:62–74.
- [95] OpenGL (2021). OpenGL, the industry’s foundation for high performance graphics. <https://www.opengl.org/>.
- [96] Pardo, X. M., Carreira, M. J., Mosquera, A., and Cabello, D. (2001). A snake for CT image segmentation integrating region and edge information. *Image and Vision Computing*, 19(7):461–475.
- [97] Patrick, J. and Indu, M. (2016). A semi-automated technique for vertebrae detection and segmentation from CT images of spine. In *2016 International Conference on Communication Systems and Networks (ComNet)*, pages 44–49. IEEE.
- [98] Pekelny, Y. and Gotsman, C. (2008). Articulated object reconstruction and markerless motion capture from depth video. In *Computer Graphics Forum*, volume 27, pages 399–408. Wiley Online Library.
- [99] Peter, R., Malinsky, M., Ourednicek, P., and Jan, J. (2013). 3D CT spine data segmentation and analysis of vertebrae bone lesions. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2376–2379. IEEE.
- [100] Poitout, D. G., editor (2002). *Bone metastases : medical, surgical and radiological treatment*. Springer, London.
- [101] Prevrhal, S., Fox, J. C., Shepherd, J. A., and Genant, H. K. (2003). Accuracy of CT-based thickness measurement of thin structures: modeling of limited spatial resolution in all three dimensions. *Medical physics*, 30(1):1–8.
- [102] Rager, O., Nkoulou, R., Exquis, N., Garibotto, V., Tabouret-Viaud, C., Zaidi, H., Amzalag, G., Lee-Felker, S. A., Zilli, T., and Ratib, O. (2017). Whole-body SPECT/CT versus planar bone scan with targeted SPECT/CT for metastatic workup. *BioMed research international*, 2017.
- [103] Rawla, P. (2019). Epidemiology of prostate cancer. *World journal of oncology*, 10(2):63.
- [104] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

- [105] Roth, H. R., Lu, L., Liu, J., Yao, J., Seff, A., Cherry, K., Kim, L., and Summers, R. M. (2015a). Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE transactions on medical imaging*, 35(5):1170–1181.
- [106] Roth, H. R., Yao, J., Lu, L., Stieger, J., Burns, J. E., and Summers, R. M. (2015b). Detection of sclerotic spine metastases via random aggregation of deep convolutional neural network classifications. In *Recent advances in computational methods and clinical applications for spine imaging*, pages 3–12. Springer.
- [107] Rubens, R. and Fogelman, I., editors (1991). *Bone metastases : diagnosis and treatment*. Springer-Verlag, London.
- [108] Sakamoto, R., Yakami, M., Fujimoto, K., Nakagomi, K., Kubo, T., Emoto, Y., Akasaka, T., Aoyama, G., Yamamoto, H., Miller, M. I., et al. (2017). Temporal subtraction of serial CT images with large deformation diffeomorphic metric mapping in the identification of bone metastases. *Radiology*, 285(2):629–639.
- [109] Sánchez, J. C. G., Magnusson, M., Sandborg, M., Tedgren, Å. C., and Malusek, A. (2020). Segmentation of bones in medical dual-energy computed tomography volumes using the 3D U-Net. *Physica Medica*, 69:241–247.
- [110] Sebastian, T. B., Tek, H., Crisco, J. J., and Kimia, B. B. (2003). Segmentation of carpal bones from CT images using skeletally coupled deformable models. *Medical Image Analysis*, 7(1):21–45.
- [111] Sebastian, T. B., Tek, H., Crisco, J. J., Wolfe, S. W., and Kimia, B. B. (1998). Segmentation of carpal bones from 3D CT images using skeletally coupled deformable models. In *International conference on medical image computing and computer-assisted intervention*, pages 1184–1194. Springer.
- [112] Seim, H., Kainmueller, D., Heller, M., Lamecker, H., Zachow, S., and Hege, H.-C. (2008). Automatic segmentation of the pelvic bones from CT data based on a statistical shape model. *VCBM*, 8:93–100.
- [113] Seyfried, T. N. and Huysentruyt, L. C. (2013). On the origin of cancer metastasis. *Critical Reviews™ in Oncogenesis*, 18(1-2).
- [114] Sim., F. H., editor (1988). *Diagnosis and management of metastatic bone disease : a multidisciplinary approach*. Raven Press, New York.
- [115] Skriniskas, T., Clemons, M., Freedman, O., Weller, I., and Whyne, C. (2009). Automated CT-based analysis to detect changes in the prevalence of lytic bone metastases from breast cancer. *Clinical & experimental metastasis*, 26(2):97–103.
- [116] Smith-Bindman, R., Miglioretti, D. L., and Larson, E. B. (2008). Rising use of diagnostic medical imaging in a large integrated health system. *Health affairs*, 27(6):1491–1502.
- [117] Snoeks, T. J., Baiker, M., Kaijzel, E. L., Lelieveldt, B. P., and Löwik, C. W. (2012). CT-based handling and analysis of preclinical multimodality imaging data of bone metastases. *BoneKEy reports*, 1.

- [118] Soeharno, H., Povegliano, L., and Choong, P. F. (2018). Multimodal treatment of bone metastasis—a surgical perspective. *Frontiers in endocrinology*, 9:518.
- [119] Sørensen, T., Sørensen, T., Sørensen, T., SORENSEN, T., Sorensen, T., Sorensen, T., and Biering-Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. *Biol Skr.*
- [120] Sorkine, O. and Alexa, M. (2007). As-rigid-as-possible surface modeling. In *Symposium on Geometry processing*, volume 4, pages 109–116.
- [121] Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190.
- [122] Suetens, P. (2017). *Fundamentals of medical imaging / Paul Suetens, Katholieke Universiteit Leuven, Belgium*. Cambridge University Press, [2017 edition]. edition.
- [123] Sugano, N., Sasama, T., Nakajima, Y., Sato, Y., Nishii, T., Iida, T., Nakagawa, K., Ono, K., Nishihara, S., Tamura, S., et al. (2001). Effects of CT threshold value to make a surface bone model on accuracy of shape-based registration in a CT-based navigation system for hip surgery. In *International Congress Series*, volume 1230, pages 319–324. Elsevier.
- [124] Sumner, D., Olson, C., Freeman, P., Lobick, J., and Andriacchi, T. (1989). Computed tomographic measurement of cortical bone geometry. *Journal of biomechanics*, 22(6-7):649–653.
- [125] Sun, L., Kong, Q., Huang, Y., Yang, J., Wang, S., Zou, R., Yin, Y., and Peng, J. (2020). Automatic segmentation and measurement on knee computerized tomography images for patellar dislocation diagnosis. *Computational and Mathematical Methods in Medicine*, 2020.
- [126] Süßmuth, J., Winter, M., and Greiner, G. (2008). Reconstructing animated meshes from time-varying point clouds. In *Computer Graphics Forum*, volume 27, pages 1469–1476. Wiley Online Library.
- [127] Suzani, A., Seitel, A., Liu, Y., Fels, S., Rohling, R. N., and Abolmaesumi, P. (2015). Fast automatic vertebrae detection and localization in pathological CT scans-a deep learning approach. In *International conference on medical image computing and computer-assisted intervention*, pages 678–686. Springer.
- [128] Taghizadeh, E., Terrier, A., Becce, F., Farron, A., and Büchler, P. (2019). Automated CT bone segmentation using statistical shape modelling and local template matching. *Computer methods in biomechanics and biomedical engineering*, 22(16):1303–1310.
- [129] Tahmoush, D. and Samet, H. (2006). Using image similarity and asymmetry to detect breast cancer. In *Medical Imaging 2006: Image Processing*, volume 6144, page 61441S. International Society for Optics and Photonics.

- [130] Tam, G. K., Cheng, Z.-Q., Lai, Y.-K., Langbein, F. C., Liu, Y., Marshall, D., Martin, R. R., Sun, X.-F., and Rosin, P. L. (2012). Registration of 3D point clouds and meshes: A survey from rigid to nonrigid. *IEEE transactions on visualization and computer graphics*, 19(7):1199–1217.
- [131] Torosdagli, N., Liberton, D. K., Verma, P., Sincan, M., Lee, J., Pattanaik, S., and Bagci, U. (2017). Robust and fully automated segmentation of mandible from CT scans. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 1209–1212. IEEE.
- [132] Toth, D. F., Töpker, M., Mayerhöfer, M. E., Rubin, G. D., Furtner, J., Asenbaum, U., Karanikas, G., Weber, M., Czerny, C., Herold, C. J., et al. (2014). Rapid detection of bone metastasis at thoracoabdominal CT: accuracy and efficiency of a new visualization algorithm. *Radiology*, 270(3):825–833.
- [133] Treece, G. (2021). Stradview. <https://mi.eng.cam.ac.uk/Main/StradView>.
- [134] Treece, G. and Gee, A. (2018). Cortical bone mapping: measurement and statistical analysis of localised skeletal changes. *Current osteoporosis reports*, 16(5):617–625.
- [135] Treece, G. M. and Gee, A. H. (2015). Independent measurement of femoral cortical thickness and cortical bone density using clinical CT. *Medical image analysis*, 20(1):249–264.
- [136] Treece, G. M., Gee, A. H., Mayhew, P., and Poole, K. E. (2010). High resolution cortical bone thickness measurement from clinical CT data. *Medical image analysis*, 14(3):276–290.
- [137] Treece, G. M., Poole, K. E., and Gee, A. H. (2012). Imaging the femoral cortex: thickness, density and mass from clinical CT. *Medical image analysis*, 16(5):952–965.
- [138] Treece, G. M., Prager, R. W., and Gee, A. H. (1999). Regularised marching tetrahedra: improved iso-surface extraction. *Computers & Graphics*, 23(4):583–598.
- [139] Treece, G. M., Prager, R. W., Gee, A. H., and Berman, L. (2000). Surface interpolation from sparse cross sections using region correspondence. *IEEE transactions on medical imaging*, 19(11):1106–1114.
- [140] Truc, P. T., Kim, T.-S., Lee, S., and Lee, Y.-K. (2010). A study on the feasibility of active contours on automatic CT bone segmentation. *Journal of digital imaging*, 23(6):793–805.
- [141] Ueno, M., Aoki, T., Murakami, S., Kim, H., Terasawa, T., Fujisaki, A., Hayashida, Y., and Korogi, Y. (2018). CT temporal subtraction method for detection of sclerotic bone metastasis in the thoracolumbar spine. *European journal of radiology*, 107:54–59.
- [142] UK, C. R. (2020). CT reconstruction example from the biomedical imaging group EPFL.
- [143] Unser, M. (2020). Prostate cancer incidence statistics.

- [144] Vaitiekūnas, M., Jegelevičius, D., Sakalauskas, A., and Grybauskas, S. (2020). Automatic method for bone segmentation in cone beam computed tomography data set. *Applied Sciences*, 10(1):236.
- [145] van Aarle, W., Batenburg, K. J., and Sijbers, J. (2011). Optimal threshold selection for segmentation of dense homogeneous objects in tomographic reconstructions. *IEEE transactions on medical imaging*, 30(4):980–989.
- [146] Wand, M., Adams, B., Ovsjanikov, M., Berner, A., Bokeloh, M., Jenke, P., Guibas, L., Seidel, H.-P., and Schilling, A. (2009). Efficient reconstruction of nonrigid shape and motion from real-time 3D scanner data. *ACM Transactions on Graphics (TOG)*, 28(2):1–15.
- [147] Wang, J., Fang, Z., Lang, N., Yuan, H., Su, M.-Y., and Baldi, P. (2017). A multi-resolution approach for spinal metastasis detection using deep siamese neural networks. *Computers in biology and medicine*, 84:137–146.
- [148] Wang, L. I., Greenspan, M., and Ellis, R. (2006). Validation of bone segmentation and improved 3-D registration using contour coherency in CT data. *IEEE transactions on medical imaging*, 25(3):324–334.
- [149] Wels, M., Kelm, B. M., Tsymbal, A., Hammon, M., Soza, G., Sühling, M., Cavallaro, A., and Comaniciu, D. (2012). Multi-stage osteolytic spinal bone lesion detection from CT data with internal sensitivity control. In *Medical Imaging 2012: Computer-Aided Diagnosis*, volume 8315, page 831513. International Society for Optics and Photonics.
- [150] Westin, C.-F., Warfield, S., Bhalerao, A., Mui, L., Richolt, J., and Kikinis, R. (1998). Tensor controlled local structure enhancement of CT images for bone segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1205–1212. Springer.
- [151] Wetzel, L. H., Levine, E., and Murphey, M. D. (1987). A comparison of MR imaging and CT in the evaluation of musculoskeletal masses. *Radiographics*, 7(5):851–874.
- [152] Whitmarsh, T., Treece, G. M., and Poole, K. E. (2014). Automatic segmentation and discrimination of connected joint bones from CT by multi-atlas registration. In *Computational Methods and Clinical Applications for Spine Imaging*, pages 199–207. Springer.
- [153] Whitted, T. (1979). An improved illumination model for shaded display. In *Proceedings of the 6th annual conference on Computer graphics and interactive techniques*, page 14.
- [154] Whyne, C., Hardisty, M., Wu, F., Skrinckas, T., Clemons, M., Gordon, L., and Basran, P. S. (2007). Quantitative characterization of metastatic disease in the spine. Part II. histogram-based analyses. *Medical physics*, 34(8):3279–3285.
- [155] Wiese, T., Burns, J., Yao, J., and Summers, R. M. (2011). Computer-aided detection of sclerotic bone metastases in the spine using watershed algorithm and support vector machines. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 152–155. IEEE.

- [156] Wiese, T., Yao, J., Burns, J. E., and Summers, R. M. (2012). Detection of sclerotic bone metastases in the spine using watershed algorithm and graph cut. In *Medical Imaging 2012: Computer-Aided Diagnosis*, volume 8315, page 831512. International Society for Optics and Photonics.
- [157] Wu, D., Liu, D., Puskas, Z., Lu, C., Wimmer, A., Tietjen, C., Soza, G., and Zhou, S. K. (2012). A learning based deformable template matching method for automatic rib centerline extraction and labeling in CT images. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 980–987. IEEE.
- [158] Xu, Y. and Goodacre, R. (2018). On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing*, 2(3):249–262.
- [159] Yang, D., Xiong, T., Xu, D., Zhou, S. K., Xu, Z., Chen, M., Park, J., Grbic, S., Tran, T. D., Chin, S. P., et al. (2017). Deep image-to-image recurrent network with shape basis learning for automatic vertebra labeling in large-scale 3D CT volumes. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 498–506. Springer.
- [160] Yang, L., Xie, Y., Li, B., Xie, M., Wang, X., and Zhang, J. (2015). Symmetry based prostate cancer detection. *The British journal of radiology*, 88(1050):20150132.
- [161] Yao, J., Burns, J. E., Sanoria, V., and Summers, R. M. (2017). Mixed spine metastasis detection through positron emission tomography/computed tomography synthesis and multiclassifier. *Journal of medical imaging*, 4(2):024504.
- [162] Yao, J., O'Connor, S. D., and Summers, R. (2006). Computer aided lytic bone metastasis detection using regular CT images. In *Medical Imaging 2006: Image Processing*, volume 6144, page 614459. International Society for Optics and Photonics.
- [163] Yao, J., O'Connor, S. D., and Summers, R. M. (2007). Computer aided detection of lytic bone metastases in the spine using routine CT images. In *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 512–515. IEEE.
- [164] Yip, S., Perk, T., and Jeraj, R. (2014). Development and evaluation of an articulated registration algorithm for human skeleton registration. *Physics in Medicine & Biology*, 59(6):1485.
- [165] Yu, H., Wang, H., Shi, Y., Xu, K., Yu, X., and Cao, Y. (2018). The segmentation of bones in pelvic CT images based on extraction of key frames. *BMC medical imaging*, 18(1):18.
- [166] Zhang, J., Yan, C.-H., Chui, C.-K., and Ong, S.-H. (2010). Fast segmentation of bone in CT images using 3D adaptive thresholding. *Computers in biology and medicine*, 40(2):231–236.
- [167] Zhang, Z. (1994). Iterative point matching for registration of free-form curves and surfaces. *International journal of computer vision*, 13(2):119–152.

-
- [168] Zhou, S., Cheng, Y., Wang, Y., Dong, K., Guo, C., Bai, J., and Tamura, S. (2013). Segmentation of the hip joint in CT volumes using adaptive thresholding classification and normal direction correction. *Journal of the Chinese Institute of Engineers*, 36(8):1059–1072.