Segmentation of Body Regions and Airways in Full-Body Digital Paediatric X-Ray Images

Murray Long (LngMur001)

August 10, 2008

Acknowledgments

UCT's Lodox programme without whose generous financial support this work would not have been possible.

Tania Douglas for putting up with all my missed deadlines.

Mum and Dad, who explained the difference between "where" and "were"!

Abstract

The objectives of the study were to develop techniques for automated segmentation of full-body digital x-ray images. Specifically, control points on the body were selected for use in segmentation of the full-body image into smaller areas of interest, and in automatic technique factor control in linear slot-scanning x-ray imaging systems. In addition, segmentation of the pulmonary airways from paediatric full-body x-ray images was considered.

For the purpose of locating the patient, algorithms based on rigid template matching, rule based methods and active appearance models where considered. An automatic system for detecting errors produced by these algorithms based on the expected shape and texture of the patient is also presented.

The set of points needed to extract the regions of interest were most reliably detected using active appearance models, which resulted in a median error of 21mm per point. The points required for the automatic technique factor control were most accurately identified using separate active appearance models for the head and torso of the patient, resulting in a median error of 5.7mm.

For airway segmentation, images where first warped onto a normalized shape using active shape models and then filtered using a local normalisation filter. Airway segmentation was attempted on the filtered images using expectation maximisation classification and snakes.

Expectation maximisation classification produced the best results for airway segmentation, however produced only a marginal improvement over selecting the most likely positions of the airways based on a set of training images, so is not expected to be of significant diagnostic value.

Contents

Chapter 1

Introduction

The Lodox Statscan x-ray imaging system (Lodox Systems, Johannesburg, South Africa) produces full-body digital x-ray images using a slot scanning geometry and is currently used primarily in the trauma setting [1]. The MRC/UCT Medical Imaging Research Unit collaborates with Lodox Systems to develop new medical applications for the system. The project presented here explores capabilities that may be included in future versions of the Statscan system.

A problem with traditional lateral spine x-rays is the large variations in width and density of the patient. When obtaining a lateral image, x-ray technique factors need to be selected to allow enough penetration through dense regions such as the shoulders but also to prevent overexposure of the film for the less dense regions over the neck and lung fields. Linear slot-scanning allows for the possibility of changing the technique factors during the scan, depending on which area is currently being scanned. This technique is known as automatic technique factor control (ATFC) and is expected to improve the image quality of lateral spine scans.

In order for ATFC to be used effectively, the positions of the patient relative to the C-arm of the x-ray machine needs to be known before the scan is started so that the technique factors can be adjusted at the correct points during the scan. Thus, the first goal of the project was to obtain, automatically, a set of control points that could be used for this purpose.

Locating the patient in a full-body x-ray image has possible applications beyond ATFC. It could be used for initializing any segmentation or computer analysis algorithm in the correct section of the image. It could be used for automatically splitting the image into various regions of interest containing specific parts of the patient's body; this would be useful for radiologists who are only interested in a certain section of the image.

It has been found that the Statscan machine allows better visualisation of the pulmonary airways than conventional x-ray imaging or computed radiography [2]. Hence it may be possible to detect the distortion of the airways caused by pulmonary tuberculosis (TB) by analysing the shape of the airways in a patient with suspected TB. Computer assisted analysis of airway shape would require the location of the lungs in the body. The full body segmentation achieved in the first part of the project was used to locate the lungs, in order to address the second goal of the project, namely segmentation of the airways for airway shape analysis.

1.1 Study Population

A set of images were obtained from the routine scans performed on trauma patients at the Red-Cross Children's Hospital. The patients were all aged between 12 and 13 years and were all admitted for non-chest related trauma. Full body radiographs of the patients at a resolution of 2.08 line pairs per mm (lp/mm) were taken using the Lodox Statscan linear slot-scanning x-ray

machine. As there was no guarantee that the patients were not suffering from a undetected pathology, every image was examined by a radiologist and excluded from the study if any evidence of pathology was found in the lungs. Images were also excluded if the patient was rotated in the transverse plane and therefore presented a distorted image of the airways. After these exclusions, a set of 83 images remained which were used in this study.

1.2 Document Structure

This document is organized into two parts. Part 1 deals with the initial problem of locating the patient and the patient's lungs within a full body x-ray. A more detailed description of this problem and an overview of relevant work done previously can be found in Chapter 2. A number of different techniques for locating the patient are considered in this section, including a rule based approach (Section 3.2), rigid template matching (Section 3.3) and active appearance models (Section 3.4). In order to improve upon some of the observed limitations of active appearance models, an approach using multiple models was considered in Section 3.5. Finally, the results from each technique are compared and the relative advantages of each discussed in Chapter 4.

Having dealt with the problem of locating the patient in Part I, segmentation of the airways is considered in Part II. As with Part I, the first chapter of Part II (Chapter 5) provides a detailed description of the problem, relevant previous work and motivations for pursuing a solution. Section 6.2 deals with preprocessing operations applied to the airway images in order to improve the visibility of the airways. Two methods of airway segmentation are then considered, namely expectation maximization classification (Section 6.3) and snakes (Section 6.4). Conclusions regarding the effectiveness of these methods for segmenting the airways as well as recommendations for the direction of further work on this problem can be found in Chapter 7.

Part I

Full Body Segmentation

Chapter 2

Background

2.1 Problem Description

The Lodox Statscan, shown in Figure 2.1, is a low-dose linear slot scanning x-ray machine primarily used to capture full body images of trauma patients. The device consists an x-ray generator and detector which are mounted on opposite ends of a C-arm. When performing a scan, the C-arm is moved down the length of the patient allowing a thin fan beam of radiation to pass though the patient and onto the detector. A full body image is assembled from the recorded amount of radiation striking the detector at each position passed during the scan. The C-arm can be rotated around the patient to allow for different views of the patient. An anterior-posterior scan (shown in Figure 2.1a) is taken with the C-arm rotated so that the generator is positioned directly above the patient. A lateral scan (Figure 2.1b) is performed with the C-arm rotated such that the generator and detector are on either side of the patient.

Figure 2.1: The Statscan machine.

(a) Anterior-posterior scan (b) Lateral scan

Figure 2.2: Images generated by the Statscan machine.

In some cases, the full body images produced by the machine may not be desired as the radiologist may only wish to view and store smaller regions of interest. The process of manually splitting the image into the desired regions of interest is a tiresome task. If the patient could be automatically identified in the image, the entire process could be automated, saving time and frustration on the part of the operator.

The ability to locate a desired region of interest is also an important first step in applying more detailed segmentation algorithms, for example, segmentation of the airways would first require a knowledge of the location of the patient's lungs.

A third possible application lies in the automatic technique factor control used in the Statscan machine. As the machine is used primarily for trauma patients, lateral scans of the spine are frequently performed to check for spinal injuries. Due to the large variations in thickness and density of the patient's body encountered when performing a lateral scan, the technique factors need to be changed at different points along the body to produce the best quality image. This is presently accomplished by manually marking control points on an anterior-posterior scan of the patient, and altering the technique factors of the lateral scan based on the position of the x-ray beam relative to these points. If these control points could be found automatically, it would improve the ease and speed with which lateral scans of the spine could be performed.

The following section describes a number of approaches to automatically locating the required points on a full body x-ray image.

2.2 Existing Solution

In order to provide the points required for automatic technique factor control in the Lodox Statscan at present, an anterior-posterior scan is performed after which the operator is required to manually mark the following control points on the resulting image.

1. Top of the head
2. Top of shoulders
3. Sternal-clavicular junction
4. Top of diaphragm
5. Hip joint

These points are marked using a mouse to position a 1 pixel wide horizontal line on the image. To fit the $2.08 lp/mm$ image onto the screen, it is down-sampled resulting in a pixel size of $2.5mm^2$. Because points positioned by the human operator using this system are clamped onto a single pixel with a width of $2.5mm$, the human positioned points will always have an uncertainty of $2.5mm$. After the markers have been positioned, the C-arm is rotated, and the patient is instructed to lift his arms above his head so as to prevent the humerus bones from shadowing the spine. The lateral scan can now be performed. The speed of the C-arm during the scan is automatically adjusted based on the position of the control points to provide the best image quality for the changing density and thickness of the patient's body.

Once a scan has been performed, the operator may wish to split the image into smaller regions of interest. In order to achieve this automatically, the following points need to be identified:

1. Top of the head
2. Bottom of the head
3. Top of the shoulders
4. Head of the humerus
5. Elbow joint
6. Wrist Joint
7. Hip joint
8. Knee joint
9. Ankle joint

2.3 Full-Body X-ray Segmentation

Segmentation of full body x-rays (into head, torso, lung-fields, arms, legs and feet) for the purpose of isolating regions of interest, was previously considered by Biagio [3]. The segmentation was accomplished by thresholding the image and scanning the thresholded image line by line, labeling each row of pixels in turn. The label to be assigned was determined by a finite state machine whose transitions were controlled by the output of two algorithms. The first algorithm detected local minima in the width of the patient (as found at the neck and ankles) and the second detected convergences at the groin and arm-pits.

This approach has three limitations: Firstly it requires that the patient be correctly thresholded from the rest of the image, as any part of the background that is incorrectly thresholded will give an unexpected shape to the patient, confusing the finite state machine and resulting in incorrect labeling. Secondly, it requires that the patient be in the expected pose. This requirement frequently leads to problems, for example, if the patient's arms are too close to their torso the "armpit" point can be lost, again resulting in incorrect labeling.

Finally, as the algorithm is applied to a thresholded image, no information about the internal structures of the patient can be considered. Due to the large variety of body shapes and poses, the overall outline of the patient's body expresses a large amount of variation unrelated to the internal structures, so estimating the position of these structures based on the outline alone, could prove to be inaccurate.

2.4 Human Body Segmentation in Other Media

Segmentation and pose estimation of human bodies have previously been considered in the field of vision-based motion capture. While the majority of commercial applications use visible markers, there have been some attempts at locating the subject by identifying the characteristic shape of the human body as done by Mori et al [4], who attempted to locate and identify the pose of a human figure in photographs. Their technique involved tessellating the image to include all sufficiently strong edges. The resulting triangle mesh was exhaustively searched to find the best fit to a simplified model of the human body. This technique is again limited to considering the outline of the human body.

2.5 General X-ray Segmentation

Although not much work has been done on the segmentation of the entire human body from an x-ray image, there is a large amount in literature devoted to the segmentation of other features, including lungs [5], clavicles [6], metacarpals [7], ribs [8] and TB lesions [9]. Those methods which may have an application in this project are outlined below.

2.5.1 Rigid Template Matching

Rigid template matching is a simple technique which involves exhaustively searching the entire image for the best match between the image data and a template of the feature which is being searched for. This technique has been used in a variety of applications, for example, by Koeslag [9] to find lesions in lung x-ray images, and by Edwards et al. [10] as a registration step for a more detailed search.

2.5.2 Rule Based Method

Biagio [3] implemented a rule based segmentation of the patient and patient's airways by locating changes in the patients width, convergences at the connections between limbs and torso, and the relatively weaker intensity of the lung fields. Li et al. [5] used a similar method to locate the lung fields, however included an algorithm to trace the edges of the lungs.

2.5.3 Edge Tracing

Li et al. [5] used a set of rules based on tracing edges in the image and analysis of the average image intensities in an attempt to locate the lung fields. A similar edge tracing method was used by Plourde et al. [8] to delineate the ribs. The method used by Li et al. would trace a single line around the edges of the lungs whereas Plourde's algorithm allowed for multiple branching edges so would create a tree structure rather than a single edge. The algorithm was applied to both edges of a single rib creating a pair of trees. The pair of branches which were the most parallel in the opposing trees were selected as the true edges of the rib. Plourde's method compared to Li's required this added complexity because the intensity of the rib edges is much smaller than that of the border between lung fields and the edge of the rib cage.

2.5.4 Snakes

Originally proposed in 1988 by Kass et al. [11], snakes or active contour models attempt to segment an image by fitting a curve to the boundary of a region of interest in the image. A energy function is defined as a measure of how well the curve fits the data. This function is usually a combination of the curvature of the snake and the second derivatives of the image in the direction of the normals to the curve. In this way the minimum value of the energy function is found when the curve lies along edges in the image but still maintains its smoothness. A snake is initialized in the image and its position iteratively improved to minimize the energy function. This technique has been successfully applied to chest x-rays to locate the edges of ribs by Yue et al [12].

2.5.5 Active Shape Models

Active shape models [13], allow a snake-like model to be constrained based on its shape thus preventing the model from fitting to any implausible shapes. The algorithm takes as an input, a set of training images which have been manually annotated. The annotation of these images is then used to build a model of the expected shape and allowed variation within the training set. The models are fitted to the image in a similar fashion to snakes, however they are restricted to shapes found in the training set. Active shape models have been applied to X–ray image segmentation by Schilham et. al [14] to detect the lung fields in chest posterior-anterior radiographs.

2.5.6 Active Appearance Models

Active appearance models were described by Edwards et al. [10] in the field of facial recognition and have since been applied to X–ray images for locating the clavicles [6] and metacarpal bones [7]. This technique requires a manually annotated set of training images, from which a model of the expected shape and texture can be built. This model can be used to generate a synthetic image based on a set of model parameters. The model parameters which best represent the features of an unseen image are found by minimizing the error between the unseen image and the synthetic images generated by the model.

This technique has had several improvements since its introduction, including searching across multiple resolutions and model adjustments to include more information about the borders of the model [6] and the inclusion of additional optimisation stages [7].

As synthetic images are built from linear combinations of the texture and shape of the training data set, any non-linear changes in the training set are not represented well by the model. This includes the rotation of limbs one would expect to find in X–ray images. The model also requires an accurate initial registration to be able to converge on the correct solution. Hence some additional processing may be needed to position the model sufficiently accurately on the image.

The following chapters expand on the application of the above methods to full body X–ray images.

Chapter 3

Methods

3.1 Chapter Summary

This chapter outlines the methods used to locate the patient in a full-body x-ray image as described in Section 2.1. The following methods were tested:

Rule-Based Method (Section 3.2) Based on those used by Biagio and Li (Section 2.3), the rule-based method attempts to locate points on the patient by considering the changes in width and density of the patient's body, finding convergence of soft tissue and tracing the edges of bones.

Rigid Template Matching (Section 3.3) attempts to locate points by comparing every section of the image to the expected texture at the point being searched for.

Active Appearance Models (Section 3.4) as described by Cootes et. al (Section 2.5.6) locates features in an image by comparing the image features to a model of their expected shape and texture.

Articulated Active Appearance Models (Section 3.4) attempted to improve the performance of active appearance models by splitting the patients into smaller sections which could be searched individually.

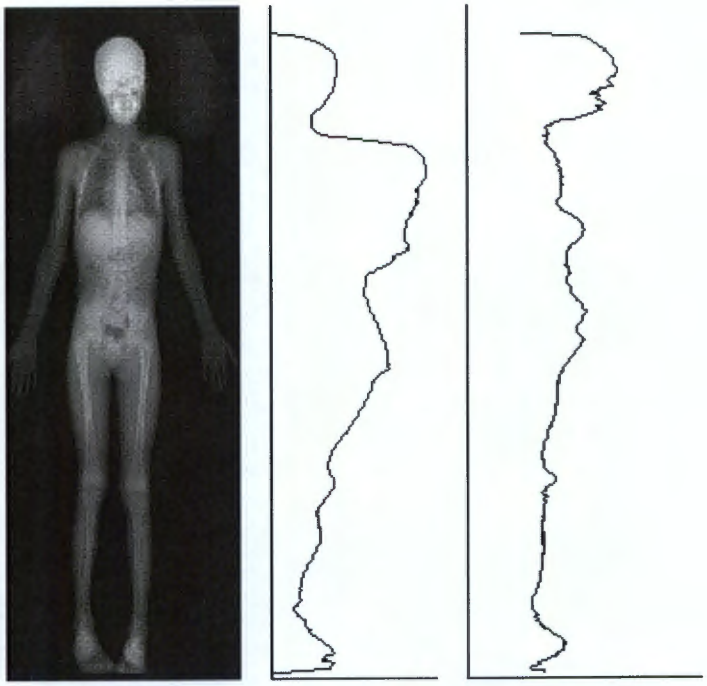
Finally, a method to detect any incorrectly identified points is described in section 3.6.

3.2 Rule-based Method

3.2.1 One Dimensional Analysis

To simplify the segmentation problem, the images were reduced to a one dimensional function that could be searched for features such as minima, maxima and points of inflection. This technique is similar to Biagio's "hour-glass algorithm" [3], which would search for local minima in the width of the patient, however it also considers local maxima and points of inflection in both the width and density functions. The one dimensional functions which were used are:

1. The width of the patient (calculated as the column-wise sum of the thresholded image)
2. The average density of the patient (calculated as the column-wise sum of the greyscale image divided by the width function)



(a) Original image (b) Width (c) Average Density

Figure 3.1: Plots of width and density functions.

As shown in Figure 3.1, the shoulders produce a sharp increase in the width function. The start of the head, diaphragm, knees and ankles produce sharp increases in the density function. By locating these features in the one dimensional function, the points can be located (in one axis) on the image. Once these points have been located in one dimension, a slice of the image in the opposing dimension can be examined to determine the position of the point in that dimension.

Thresholding

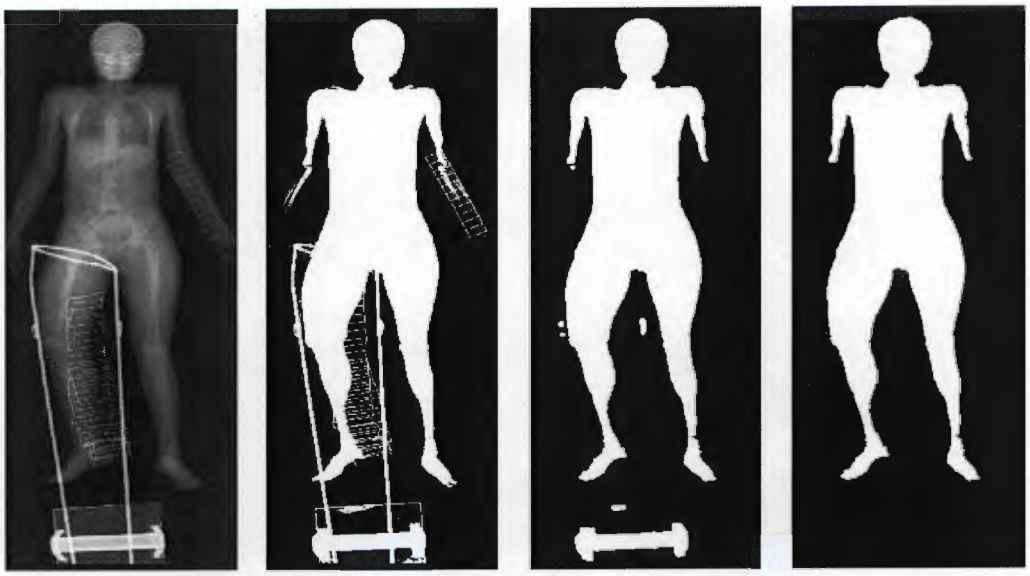
Before the width and density functions are calculated, the patient's body needs to be isolated from the rest of the image. This was accomplished using the methods suggested by Biagio [3]. The image was first split into background and foreground components using Otsu's method [15].

Otsu's method finds the threshold point, t which minimizes the within-class variance, $\sigma(t)$, for the two classes of pixels present in the image, one class with intensity greater than t and one with intensities lower than t. In our application the class with lower intensity represents the background while the class with higher intensity represents the patient.

$$\sigma(t) = n_1(t)\sigma_1(t) + n_2(t)\sigma_2(t)$$

where $\sigma_1(t)$ and $\sigma_2(t)$ represent the variance for the two classes separated by the threshold t and $n_1(t)$, $n_2(t)$ represent the number of pixels in each class.

The results of the thresholded image can be seen in Figure 3.2(b). In this image and many others, the thresholding incorrectly identifies high intensity background objects as being part of the patient. Such background objects include oxygen bottles, stretchers and braces for broken limbs (as shown in the figure).



(a) Original Image (b) Thresholding (c) Opening (d) Largest Region

Figure 3.2: Segmenting the patient.

In many of the cases, including Figure 3.2, the background objects are much thinner than the patient's body. This allows the use of morphological operations to break the connections between patient and background objects.

In reference to Figure 3.3, morphological operators [15] compare an object in a binary image, O, to a *Structuring Element*, S with a reference point R. The two basic operations which can be applied are erosion and dilation.

Erosion returns the set of possible positions for the structuring element's reference point, for which the structuring element is completely contained within the object.

Dilation returns the set of possible positions for the structuring element's reference point, which results in an overlap between the object and structuring element.

Two further operations are defined, opening (an erosion operation followed by a dilation) and closing (a dilation operation followed by an erosion).

The effects of these operations applied to a binary image, are shown in Figure 3.3. As is evident in the figure, the opening operation tends to break narrow links between objects in the image. When applied to the thresholded image of the patient, a circular structuring element with a 60mm radius was found to remove most of the background objects while preserving the shape of the patient. The results of this operation are shown in Figure 3.2(c).

Once the opening operation had been applied to break the links between patient and background structures, the size of each contiguous region of thresholded pixels was calculated. In all the images tested, the pixels representing the patient created the largest region, so all smaller, unconnected regions were deleted (Figure 3.2(d)). Once the pixels which represent the patient have been isolated in this manner, further processing can be done to locate specific points within the patient.

3.2.2 Convergence Detection

To locate the points of convergence on the patient's body such as at the groin and armpits, Biagio's algorithm would scan upwards though the patient's body, counting the number of

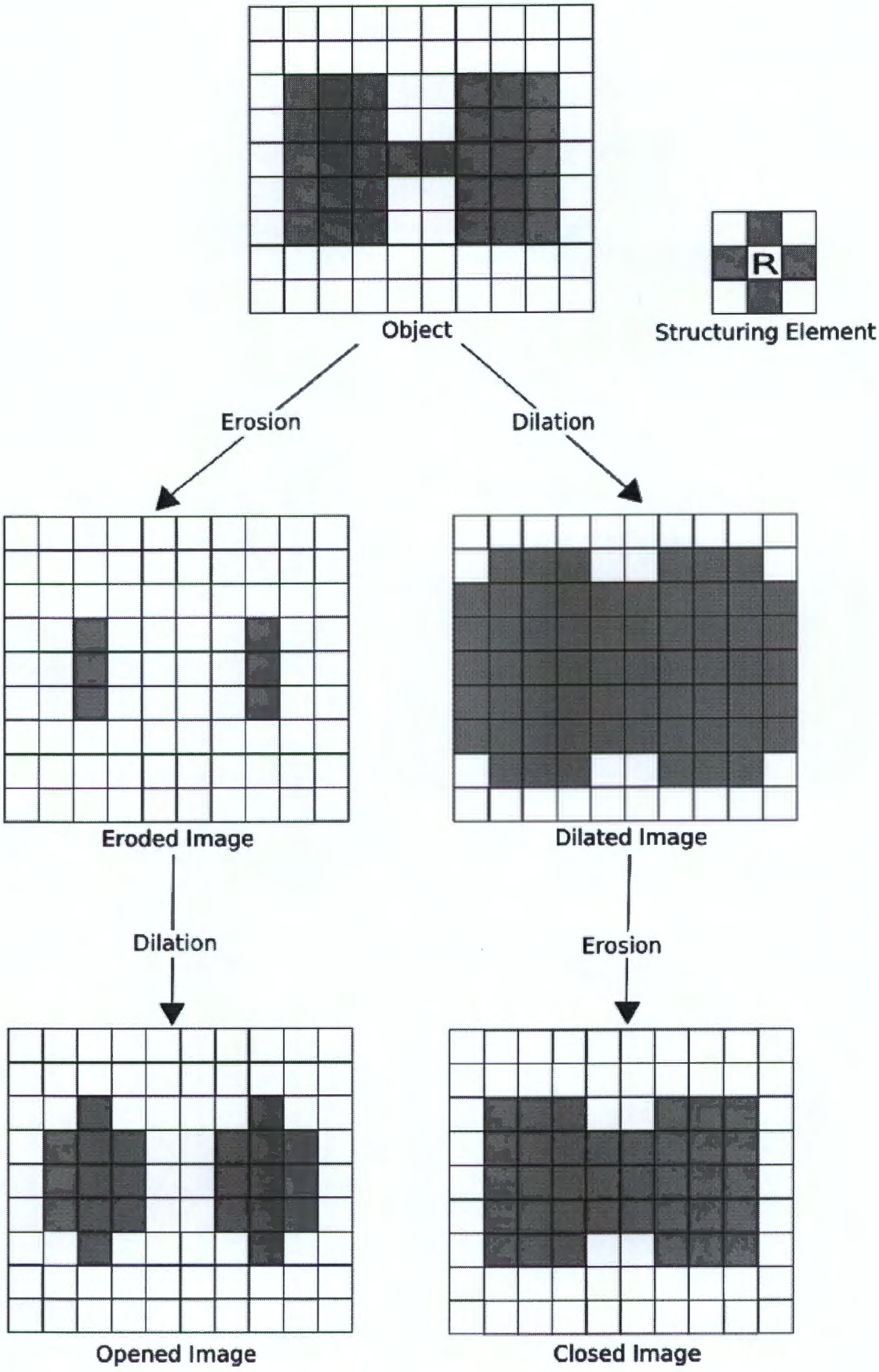


Figure 3.3: Applying morphological operations.

distinct regions in each horizontal slice of the image. Whenever the number of regions decreased, it was assumed a convergence between the two objects had occurred. This algorithm would successfully find the points of convergance, however was relatively slow to run on high resolution images.

In this application, in order to improve the execution speed of this search for the high resolution images, a search algorithm based on Newton's method was implemented. The algorithm would scan upwards from the bottom of the image, examining the number of sets of connected pixels present in each scan line. Whenever two or more of these were found, the distance between each set would be measured. By calculating the rate of change of this distance over several slices, the point of convergence between the two sets can be predicted.

The following iterative algorithm was used to find points of convergence:

1. Sample the distance between the body parts across two consecutive scan lines.
2. Calculate the rate of change of distance between parts.
3. Calculate the expected points of convergence based on rate of change of distance.
4. If this point is already closed, move to half the distance from the last point.
5. Repeat until the point of convergence is found.

3.2.3 Bone Tracing

Basing the segmentation on the shape of the thresholded image alone has several disadvantages. Due to the large range of weight, body shape and pose of the patients, any landmark based on soft tissue could possibly be inaccurate when compared to landmarks positioned on the bones of the patient. Points which do not occur on the border of the patient (such as the sternal-clavicular notch) would have to have their positions estimated from the position of points of the edge of the patient, so would be found less accurately than in a search based on the internal structures.

In an attempt to combat these problems, the long bones in the patient's limbs were found by tracing the high intensity edges of the bones themselves. This was accomplished using the edge tracing algorithm used by Li et al. [5] to locate the edges of the lung fields.

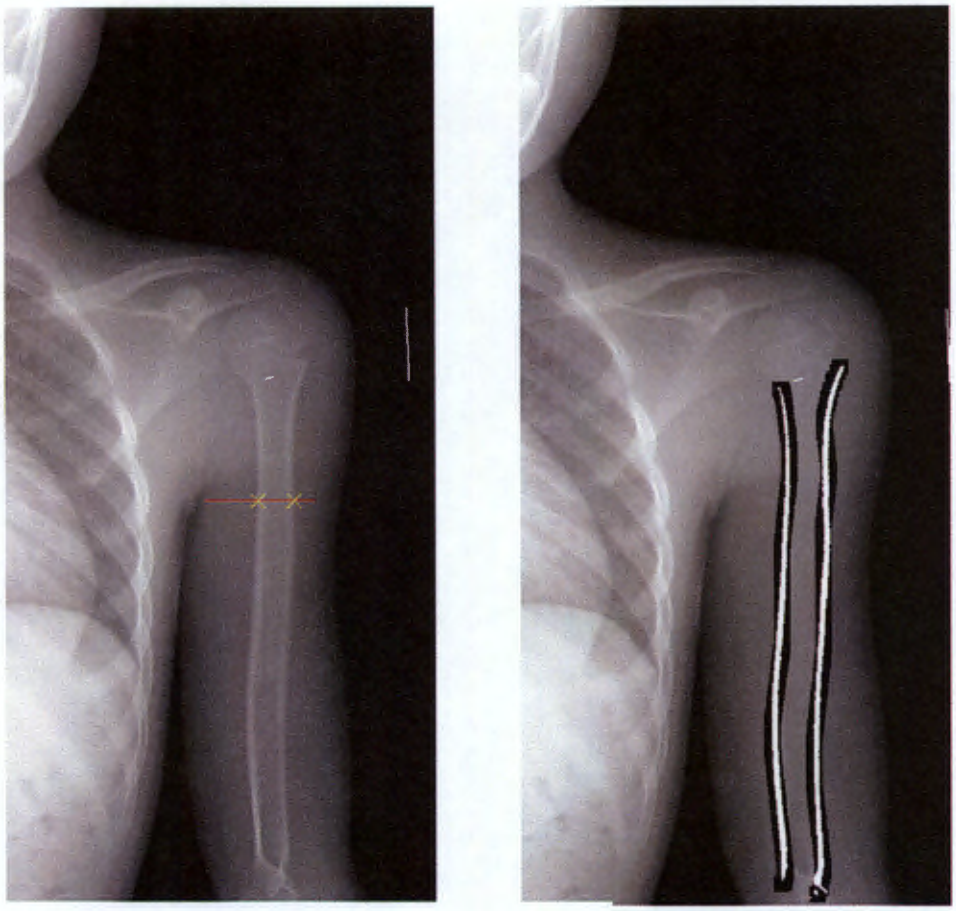
The algorithm is initialized with a point on the edge to be traced, and the approximate direction in which the edge is expected to run. In this application, such a point is found by projecting a ray (drawn in red in Figure 3.4(a)) directly outwards from the "armpit" landmark points (found by the previous algorithm) to the edge of the patient's body. The sharpest edges on this ray were assumed to be the transition between soft tissue and the denser bone of the limb (marked in yellow on Figure 3.4(a)). The expected direction of the edges are initialized to be straight up and down the length of the patient.

Once an initialization point has been found, the edge can be followed. All pixels within a given radius from the initialization point are examined, Any points which do not have sufficiently strong edges or would cause the edge tracing to exceed its maximum allowed curvature are discarded. If there are no surviving points the algorithm terminates; if there are multiple surviving points, the process is repeated on the point which requires the least amount of curvature in the the edge tracing. The direction is reset to the direction between the previous and new points.

In this manner, the edges are followed until they reach the areas of high curvature found at the ends of the bone. To provide redundancy, the edges of both sides of the bone were followed. Once both tracings have terminated, the distance travelled from the initialization point is calculated. If the distance traced for both sides of the bone is similar, the tracings are

considered to have terminated correctly, and the point for the end of the bone is positioned halfway between the end points of the two tracings. Dissimilar end points usually occur for one of two reasons; the tracing algorithm may have terminated prematurely due to unusually weak edges, high curvature or discontinuous edges (usually associated with broken bones), or, the algorithm may have jumped onto an unrelated edge (such as part of a stretcher or a support for broken bones). In order to account for both cases, the tracing which travelled the furthest is considered to be correct provided it has not travelled beyond a threshold value determined by the position of points already placed. For example, if an edge of the humerus is traced above the previously calculated position of the shoulders, it is considered to be incorrect, in which case the shorter of the two tracings is used to estimate the end of the bone.

The edges traced using this method are shown in Figure 3.4(b). This method was used to locate the extremities of all long bones in the patient's body.



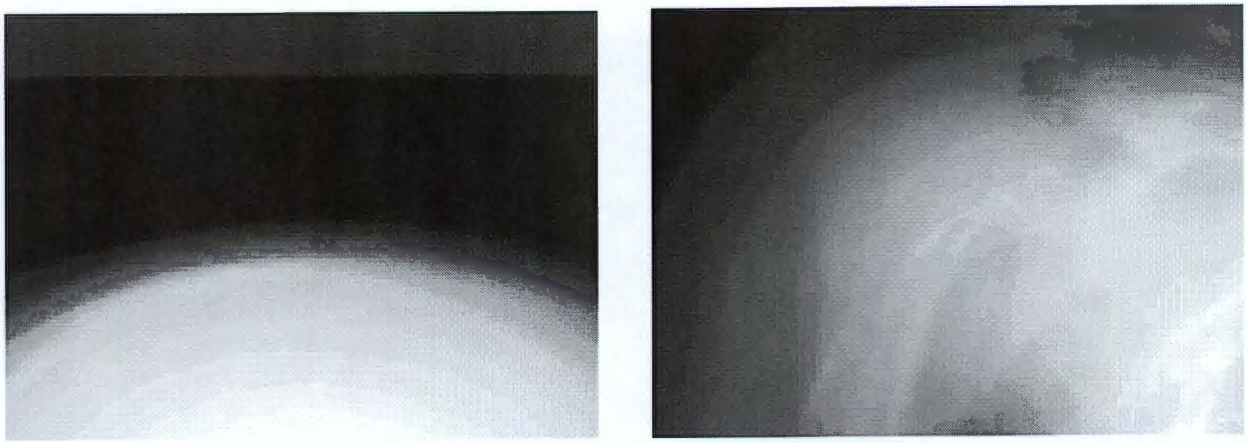
(a) Locating a point on the edge (b) Output of the tracing algorithm

Figure 3.4: Tracing the edge of the humerus.

3.3 Rigid Template Matching

On a set of 13 training images, landmark points were manually positioned by an expert. The mean expected texture in the local regions around each point was calculated by extracting a 100-pixel square region around each and finding the average of these local textures across the

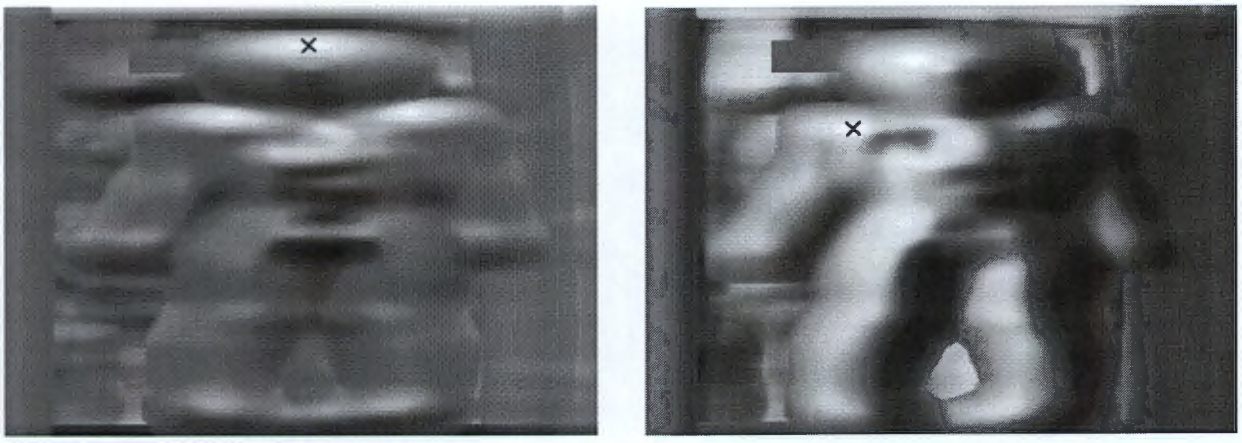
set of training images. Figure 3.5 shows the average greyscale values for the regions surrounding the top of the head and the shoulder joint.



(a) Top of The Head (b) Head of the Humerus

Figure 3.5: The average local texture around two landmark points.

When presented with an unseen image, normalized cross correlation between the unseen image and each template was used to attempt to locate the landmark points. The maximum value of the correlation represents the best match between template and image, so was considered to be the correct position for the point. Figure 3.6 show the results of a normalized cross



(a) Top of The Head (b) Head of the Humerus

Figure 3.6: The output of a normalized cross correlation between the templates shown in Figure 3.5 and a search image.

correlation between the image to be searched and the head and shoulder templates respectively. The blue crosses mark the maximum point of the cross correlation function which was deemed to be the best fit between the template and the image.

3.4 Active Appearance Models

3.4.1 Training Set

A working set of 40 images was divided at random into 13 images to be used as a training set and 27 images to be used to test the accuracy of the search. In an attempt to minimize the effects of background features on the model, all non-patient objects visible in the training image were digitally removed (as shown in Figure 3.7). Both the original and the edited images were used to build models and the results from each were compared.

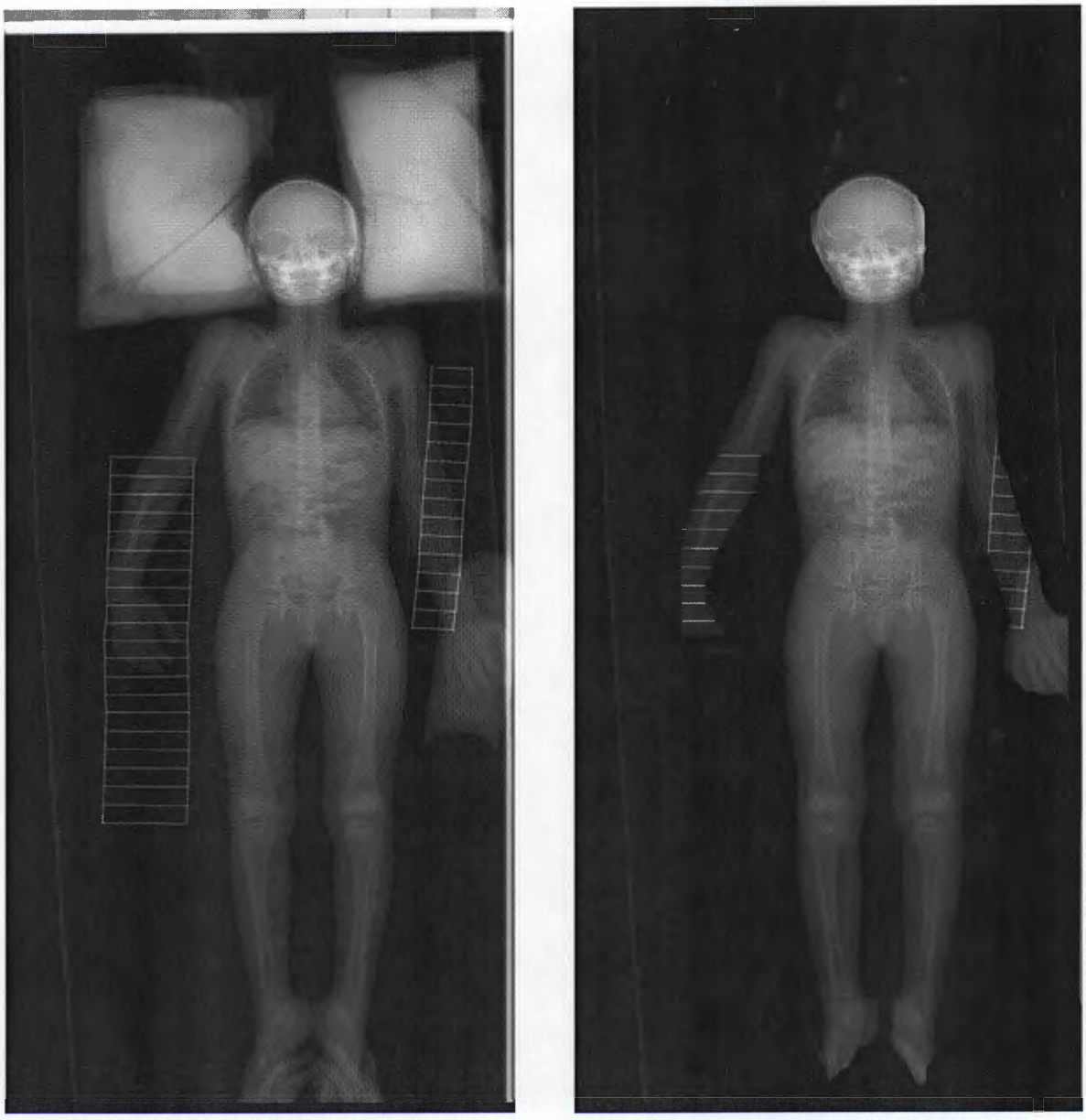
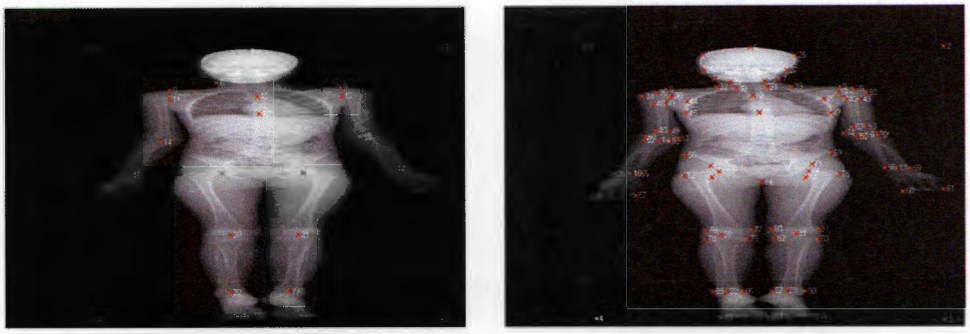


Figure 3.7: Original image (left) and the same image (right) with the background objects removed.

On all the training images, the 19 points being searched for and 4 points on each corner of the image were manually positioned. A list of these points can be found in the Appendix. The

4 points on the corners of each image were included so that the non-patient areas of the image could be tessellated. In order to evaluate the effects of the complexity of the model on the effectiveness of the search, a further 73 points were marked on the training image. These points were chosen to be easily and reliably positioned (usually at the intersection of two bones) and to be in areas where a large amount of variation between patients was evident. A full listing of the points used can be found in the Appendix.



(a) An image annotated with the 19 required points and 4 image corners

(b) The more detailed, 92 point annotation

Figure 3.8: Manual annotations of a training image.

3.4.2 Building the Model

In order to build a model representing the patient in the image, three separate components are required:

1. The position of the patient in the image;
2. the shape of the patient;
3. the greyscale texture associated with the patient.

Each of these components was concatenated into a single feature vector which would represent any instance of the model. The following sections describe how each component of this feature vector is calculated.

Position Component

In order to record the position of the patient in the image, the transform used to map the model onto the image was expressed as a 3x3 homogeneous coordinate matrix. This allows for the model to be moved, rotated, scaled and sheared in the image without effecting the representation of shape and texture. This 3x3 matrix was columnated and appended to the model feature vector; details of how initial estimates for these values were obtained can be found in Section 3.4.3.

Shape Component

The shape of every patient in the set of annotated training images was normalized in terms of rotation, position and scale and represented as the difference from the mean shape in the training set. This was accomplished using Procrustes Analysis [16], as follows.

Before two or more shapes can be compared, the effects of pose (the scale, rotation and position of the object in the image) have to be removed. Any object in an image, a, represented as a set of ordered two dimensional points (Figure 3.9(a)):

$$a = \begin{pmatrix} x_{a1} & y_{a1} \\ x_{a2} & x_{a2} \\ \vdots & \vdots \\ x_{an} & y_{an} \end{pmatrix}$$

The effects of position and scale can be removed by centering the shape on the origin and normalizing the size of the shape (Figure 3.9(b) and 3.9(c)):

$$a_{norm} = (a - (\bar{x}, \bar{y}))/S(a)$$

where $(\bar{x}, \bar{y})$ is the centroid of the shape defined as:

$$(\bar{x}, \bar{y}) = (\frac{1}{n}\sum_{j=1}^{n} x_j, \frac{1}{n}\sum_{j=1}^{n} y_j)$$

and $S(a)$ is the centroid size of a, defined as:

$$S(a) = \sum_{j=1}^{n} \sqrt{(x_j - \bar{x})^2 + (y_j - \bar{y})^2}$$

Given two position and scale normalized shapes, a_{norm} and b_{norm}, the rotational components that separate the two can be removed as follows (Figure 3.9(d)): A singular value decomposition is applied to $a^T b$,

$$UDV^T = a^T b$$

where D is a diagonal matrix whose elements represent the singular values of $a^T b$ and U and V are two unitary matrices. The transform VU^T will rotate a to optimally align with b, thereby removing the rotational component.

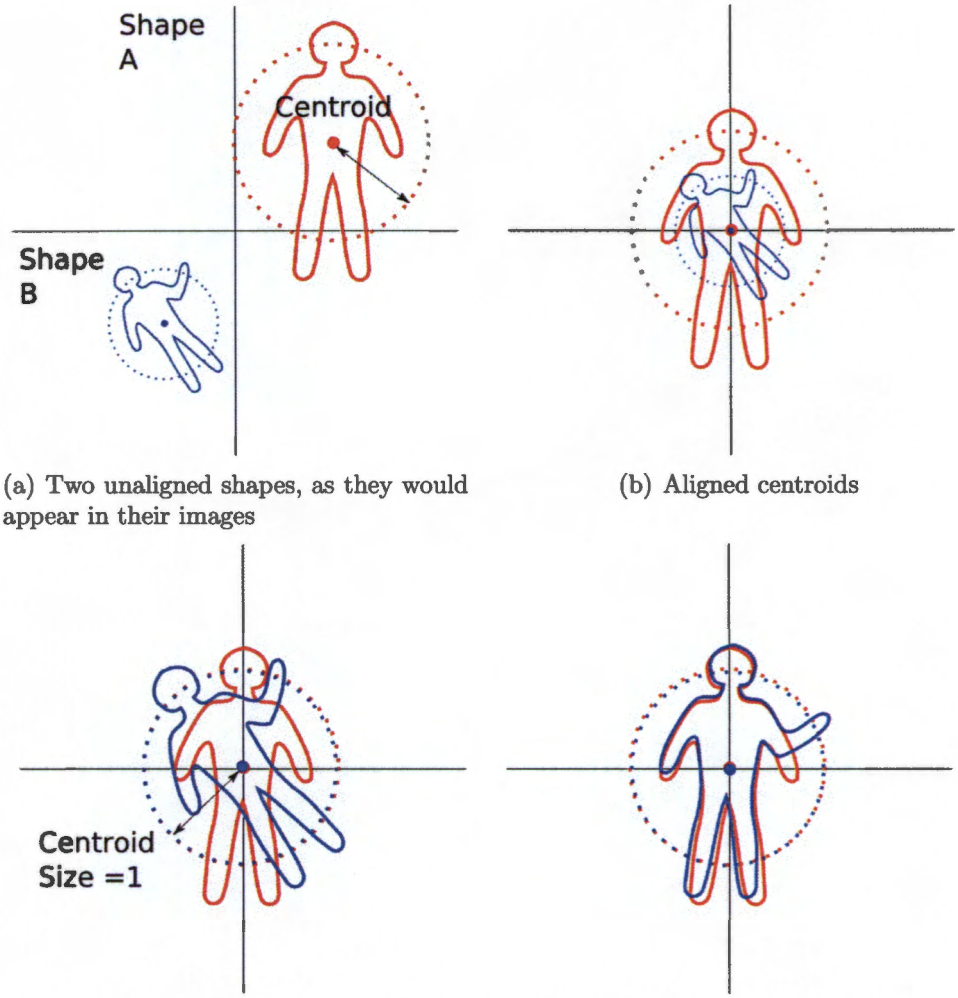
The "Procrustes Distance", P can now be used as a measure of the difference between the shapes a and b. It is calculated as:

$$P^2(a, b) = \sum_{n}^{j=1} (x_{aj} - x_{bj})^2 + (y_{aj} - y_{bj})^2$$

To determine the expected shape and allowed variation in the manually annotated training set, the mean shape is first estimated as follows:

1. On the first iteration, set the mean shape equal to the first shape in the set.
2. Align all shapes to the mean to minimize a distance metric.
3. Calculate the mean shape of all the aligned shapes.

This process is repeated until it converges on a stable mean shape. The shape component for each instance in the training set can now be calculated by subtracting the mean shape from the position, scale and rotation normalized shape for that instance.



(a) Two unaligned shapes, as they would appear in their images

(b) Aligned centroids

(c) Normalized centroid size

(d) Aligned rotation

Figure 3.9: Aligning two shapes to remove the effects of scale, translation and rotation.

Texture Component

Having obtained the mean shape, the mean texture can be calculated. The textures of each training image are warped onto the mean shape using a piece-wise affine transformation based on the manually positioned control points. The average shape-independent texture can be calculated as the average of all warped training images. This average texture is shown in Figure 3.10.

Every training image can now be represented by the (very large) vector

$$(x_1..x_n, y_1..y_n, t_1..t_r)$$

where $(t_1..t_r)$ are the greyscale intensities of the pixels which make up the texture of the warped image and $(x_1..x_n, y_1..y_n)$ are the points which define the shape of the model.

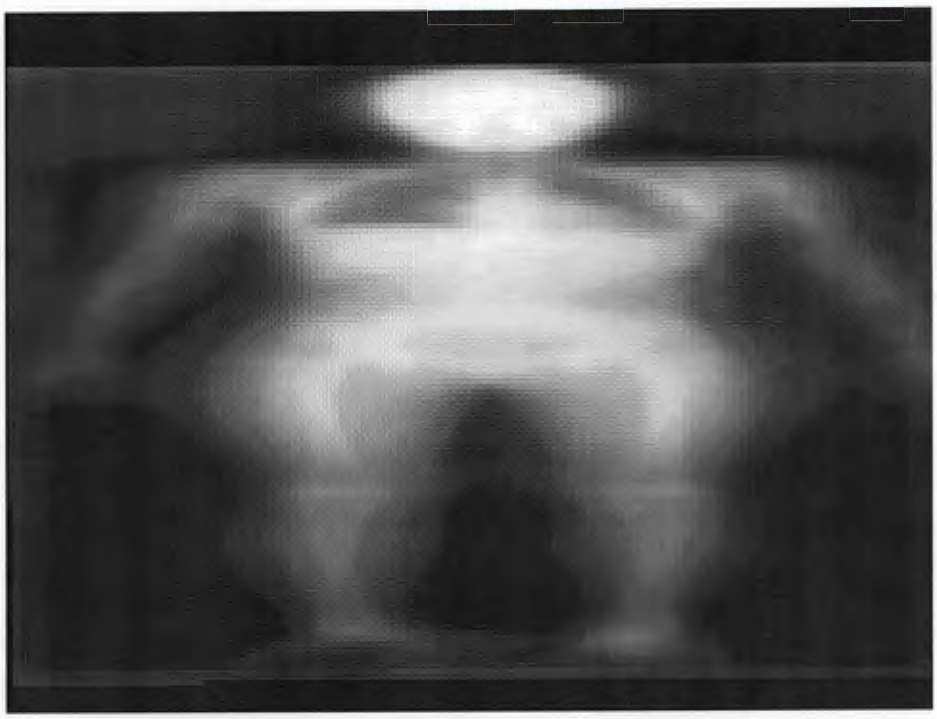
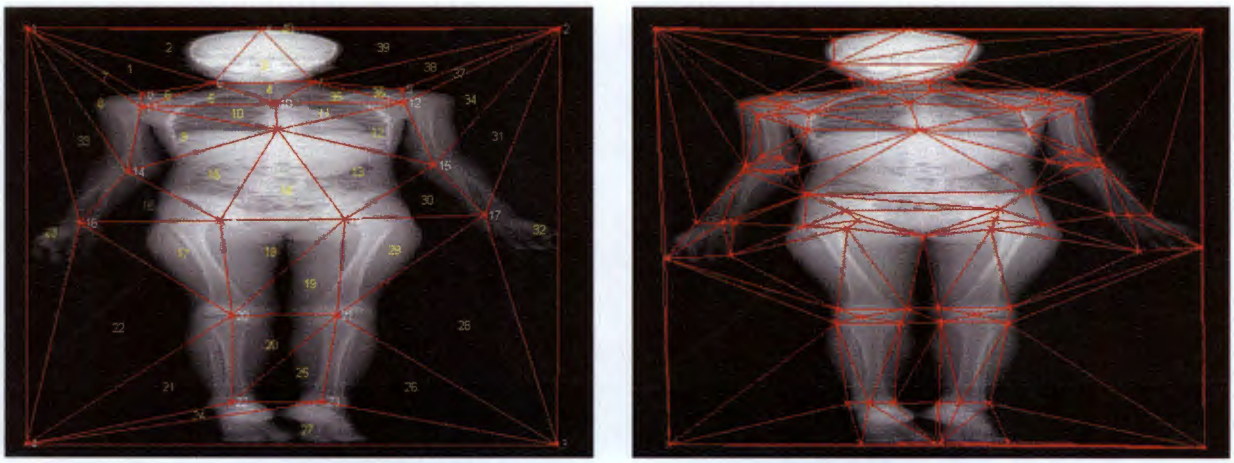


Figure 3.10: The mean texture calculated when building the appearance model.

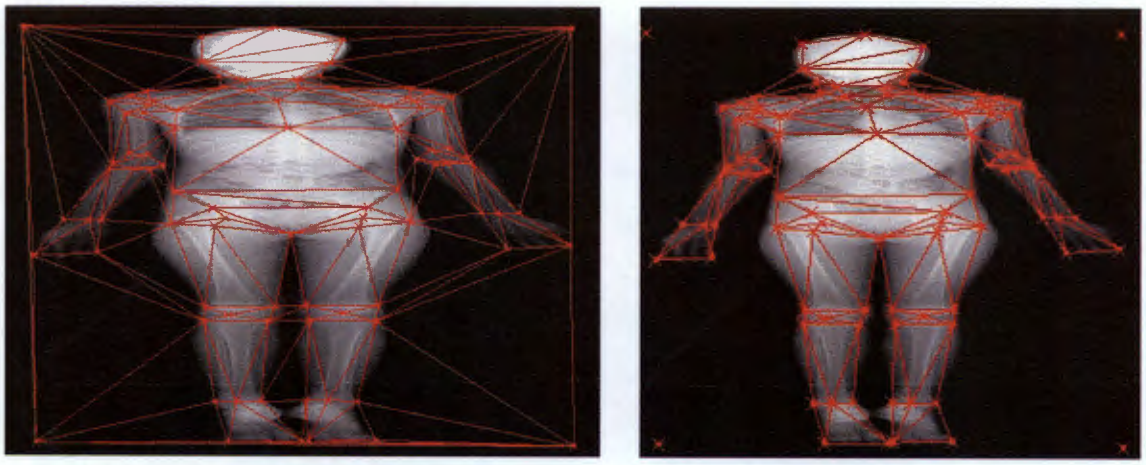
Tessellation

In order to apply the piece-wise affine transformation to the images, the point set first needs to be tessellated into a triangle mesh. Each triangle has a transform applied to it independently after which they are reassembled to form the warped image. Initially an automatic Delaunay tessellation [17] was used (the output of this algorithm is shown in Figure 3.11). However, the large triangles which span the space between arm and torso (shown in Figure 3.11(a)) can result in a blurring of model when warping the image. In order to counter this problem, the points were manually tessellated to provide the best separation between triangles within the body and triangles in the background (the manual tessellation of the 92 points model is show in Figure 3.12(a)). Given that triangles were now separated into those that were within the body and those that were not, a third tessellation method was considered, in which, in an attempt to reduce the effect of background objects, all triangles which were not part of the patient's body were deleted (shown in Figure 3.12(b)).



(a) 23 Point Model (b) 92 Point Model

Figure 3.11: Delaunay tessellation of simple and detailed models



(a) Background Triangles Preserved (b) Background Triangles Removed

Figure 3.12: Manual tessellation of the 92 point model.

Principle Component Analysis

At this stage, the model feature vectors are very long, typically containing more than 65000 elements, 9 to represent the registration of the model, 184 for the shape model, and 256^2 for a texture model at a resolution of 256x256. Searching across such a large search space is infeasible so principle component analysis was used to reduce the length of the model feature vectors and therefore the search space of the optimisation problem.

Given a set of n model vectors of m length concatenated into an $n * m$ matrix

$$x = \begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,m} \end{pmatrix}$$

Principle component analysis provides a linear transform M such that transformation of x by M

$$y = Mx$$

produces y, the transformation of x onto a new coordinate system so that the major variation in the columns of x are expressed in the first rows of y. When optimizing the model to a image, the search was performed in this new coordinate system so that the rows of y which contribute no variation to the training set can be ignored. These values are dropped from the feature vector. However when converting back to the original coordinate system, they are replaced with zero-value entries to create the original length vector.

3.4.3 Searching

The steps required to search an image using an active appearance model, can be summarized as follows:

1. Obtain an initial estimate of the model's position on the test image.
2. Based on the model's shape and position parameters, warp the image onto the mean shape.
3. Generate a synthetic image based on the model's texture parameters.
4. Calculate the error between the warped image, and the synthetic texture.
5. Based on this error, update the model parameters to improve the fit between the warped image and the synthetic texture.
6. Repeat steps 2-4 until the error function converges on a minimum value.

Registration

As documented by Ginneken et al. [6], in order for the model to converge on the correct solution it first needs to be positioned on the image fairly accurately. The required accuracy of this initial registration is dependent on the relationship between the greyscale error function and the parameter error function. If this initial registration is not accurate enough, the model will be positioned outside the range for which there is a linear relationship between these functions so the model will diverge.

Three different methods of registering the model to the image were considered:

Basic Registration: the scale and position of the model are set to the center of mass and size of the thresholded patient found in Section 3.2.1.

Rule-Based: The position, rotation, scale and shape of the model are set using the points found in the rule based system (Section 3.2).

Manually Positioned Points: The position, rotation, scale and shape of the model are set using the points manually assigned by a human operator. This is obviously not a viable method for a completely automated solution, but was used as a reference by which to compare the other two registration techniques, and to examine how the model performs under ideal conditions.

The performance of the search algorithm was compared across all three initialization methods.

Optimisation

Active appearance models provide three different error metrics to evaluate how effectively a model fits a search image.

Image error or grey-scale error: The mean-squared-per-pixel error between the synthetic image generated by the model and the image being searched.

Parameter error: The error between the parameters of the model and the ideal parameters which provide the best fit between the model and the search image.

Point error: The error between the points positioned by the model and the points on the search image.

When optimizing the model to best fit the image, the goal is to minimize the point error, thus positioning the landmark points generated by the model as close as possible to their correct locations. However, at the time when a search is performed, the only error function available is the image error (the error between the image generated by the model and the image we are searching). The relationship between the image error and the parameter error can be estimated from the training set. Assuming a linear relationship between the image and parameter errors, a Jacobian matrix is built from the training set using multivariate linear regression. Parameter offsets of 0.5 standard deviations of each model parameter are used as suggested by Ginneken et al. [6]. The parameter error can now be estimated as

$$\delta\overline{y} = A\delta I$$

where $\delta\overline{y}$ is the parameter error, A is the Jacobian matrix and δI is the greyscale error. Unfortunately the relationship between greyscale error and parameter error is not linear, so this relationship breaks down for large error values. In order to improve the accuracy, the search process is repeated over several small iterations until the greyscale error converges to a minimum value.

With reference to Figure 3.13, the AAM search can be summarised as follows: The initial registration provides the initial registration parameters for the model; these parameters are combined with the mean shape and texture values to create the first model feature vector. This feature vector is then transformed into the principle component space using the transformation matrix built in Section 3.4.2; the parameters which do not contribute to the variation in the training set are then discarded. In order to convert back to the original feature vector, the discarded components are replaced with zero values and transformed back into the original coordinate system. The registration parameters of the model are extracted from the feature vector and assembled into a homogeneous coordinate transform matrix which is used to transform the shape parameters of the model into a set of points in image space representing the

shape of the patient. The point set and pre-calculated tessellation are used to warp the search image, creating the shape-normalised search image (Section 3.4.2).

The texture parameters in the model are extracted from the feature vector and assembled into the synthetic texture. The image error is calculated as the difference between the shape-normalised search image and the synthetic texture. The image error is columnated and multiplied by the Jacobian matrix to estimate the parameter error.

The estimated parameter error is used to update the current best estimate of the model parameters; the updated parameters are then transformed using the inverse of the principle component analysis transform matrix (Section 3.4.2), to create the next iteration of the model.

This process is repeated until two iterations have passed without improvement to the image error. At this stage, the search is considered to have converged so the predicted shape is calculated one final time and presented as the results of the search. Figure 3.14 shows the convergence of the error function during this process.

3.5 Articulated Active Appearance Models

In order to better represent the expected range of motion of the patient's limbs, an articulated hierarchy of active appearance models was used with each model representing one of the following parts of the patient's body:

- Head
- Torso
- Upper arm
- Upper leg
- Lower arm
- Lower leg

Each model was assembled into the hierarchy shown in Figure 3.5. Any single model in the hierarchy will have another model to which it is attached higher up in the hierarchy, called the "parent". The model may also have one or more "children" (models attached to this one lower down in the hierarchy). Each model's position in the hierarchy determines at what stage it is optimized; parent models are optimized first, followed by their children. The points of attachment between models (shown as yellow crosses in the figure) define the axis about which a child model may rotate during its registration.

Searching

When searching a new image, the torso model was registered and optimized using the same technique as for active appearance models (Section 3.4). Following this, any changes to scale effected during the search would be propagated to each of the child models attached to the torso. The positions of each child were then updated based on the location of the appropriate attachment node on the torso. The initial rotation of the child was determined by exhaustively searching though the full range of rotation of the limb about the point of the attachment and selecting the configuration which resulted in the least error between the model and search image. The child model would the be optimized and the process repeated on its children until every model in the hierarchy had been processed.

A possible problem with this method would be the propagation of errors though the hierarchy, however by reducing the amount of variation needed to be expressed by each model, better

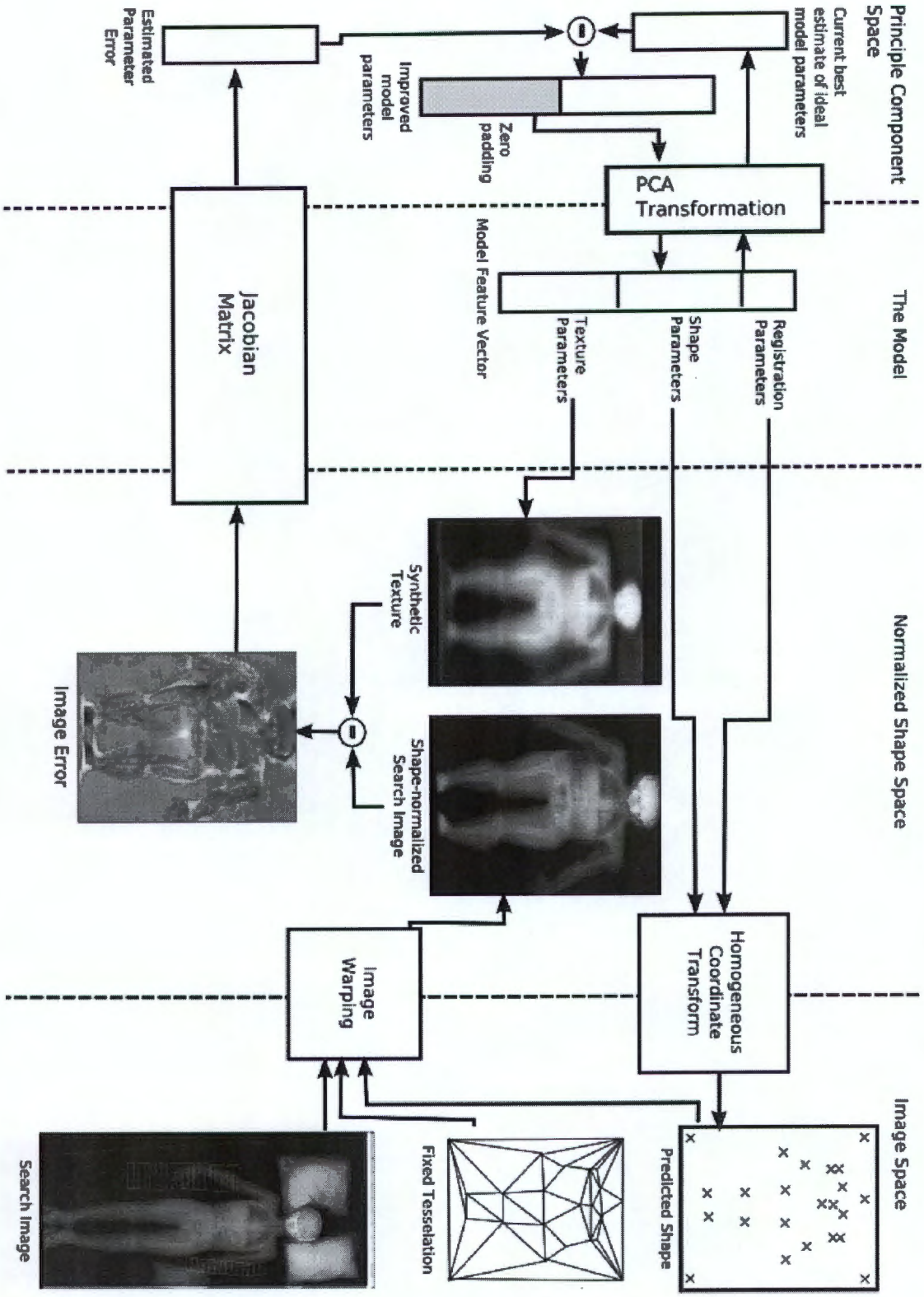


Figure 3.13: AAM Search

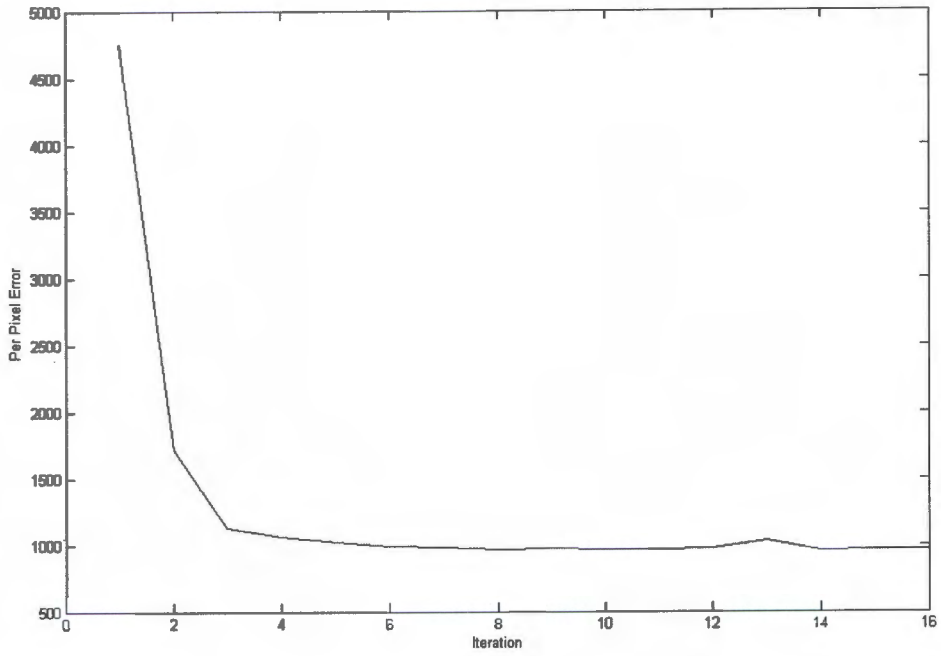


Figure 3.14: Minimization of the greyscale error during model optimisation.

matches could be obtained. This method also takes advantage of the symmetrical nature of the human body by using the same model for left and right limbs which effectively doubles the size of the training set for the repeated models.

3.6 Incorrect Point Detection

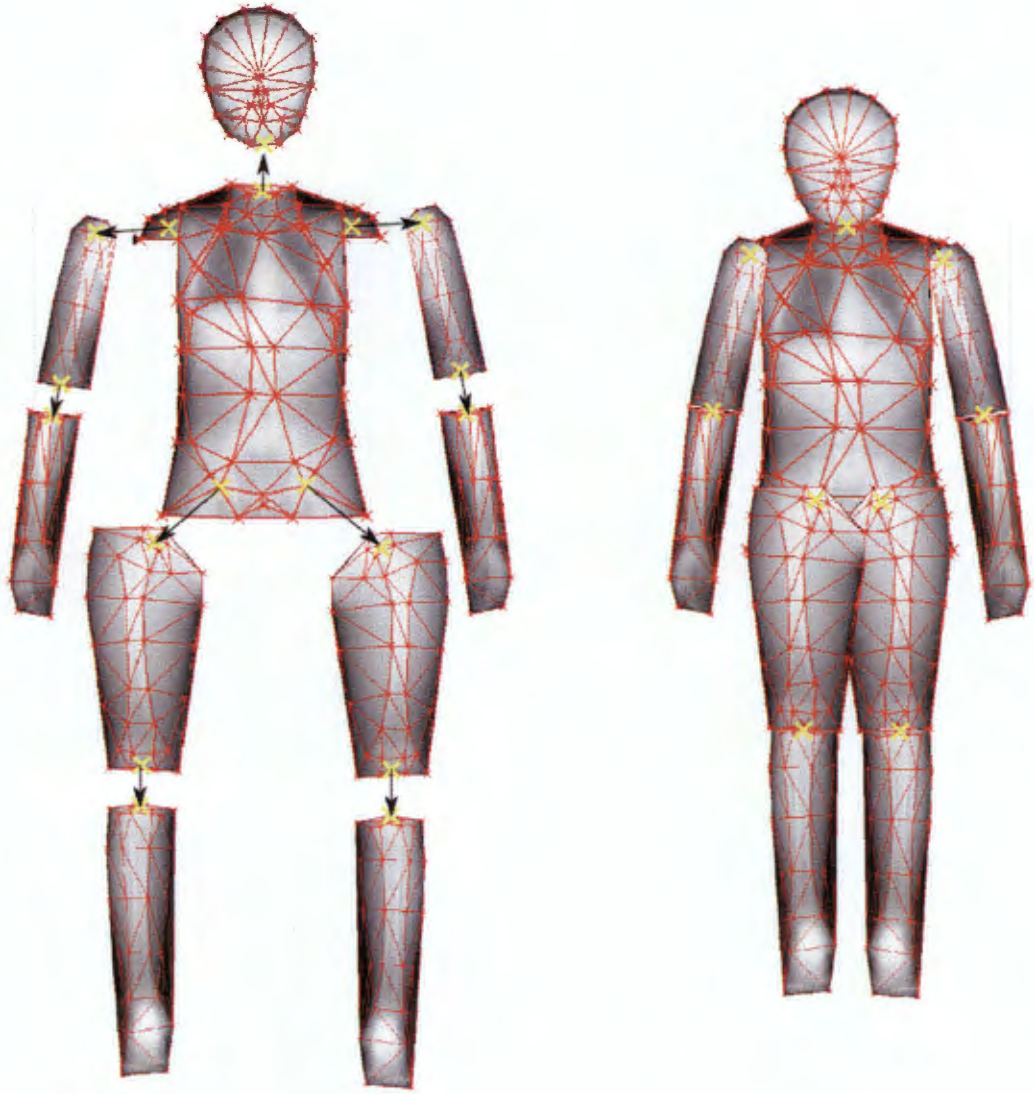
In the event that the segmentation fails, it would be useful to be able to identify the incorrectly positioned points so that the user could be alerted to the fact, or so that additional processing could be performed in an attempt to correct them. Two methods to accomplish this were considered, based on the patient's expected shape and greyscale texture respectively.

3.6.1 Shape

As most patients are expected to have similar shapes, the occurrence of incorrectly positioned points could be detected by the change in the expected shape induced by the error.

A shape generated by the rule-based system, was first aligned with the mean shape using Procrustes analysis, then transformed into the principle component space (see Section 3.4.2). Once in principle component space, any components of the shape found to be greater than two standard deviations away from the mean value (as calculated from the training set), were clamped so that they fell within the allowed deviation. Altering a component in this fashion will affect all points in Euclidean space, however it is expected to have a greater effect on the incorrectly positioned points.

The altered shape was then transformed back into Euclidean space, and each point was compared to its original. The amount that each point had been moved due to the clamping in principle component space was used as a measure of the likelihood of that point being



(a) An exploded view of each segment.

(b) The assembled model

Figure 3.15: The hierarchy of separate models. Yellow crosses represent the points of attachment, arrows represent the direction of the parent-child relationship.

incorrectly positioned. Points that were moved more than a threshold distance were labeled as incorrect; the selection of an appropriate threshold is covered in Section 4.7.1.

This error detection method will not be useful on the active appearance model based searches as the model is already restricted to matching shapes found in the training set.

3.6.2 Texture

Active appearance models (Section 3.4), provide a model of the expected greyscale values of the image. By comparing the greyscale values around each point identified by one of the other techniques, errors could possibly be detected by a poor match between the expected and existing greyscale values in the vicinity of the incorrect points. In practice, the mean greyscale texture was calculated and warped onto the new shape (Section 3.4.2). As the images were already tessellated to allow for the warping, the greyscale error between warped mean and the original image could be calculated for each triangle in the tessellation. The error associated with each point was calculated as the per-pixel average of the error across each triangle which contained the point as a vertex. If this error was above a threshold, the point was labeled as incorrect.

Chapter 4

Results and Conclusions

4.1 Chapter Summary

This chapter describes the performance of the techniques described in Chapter 3, when applied to the problem of locating the patient in a full-body x-ray. The results of each individual technique can be found in Sections 4.3 though 4.7. Comparisons between the techniques are made in Section 4.8 and recommendations for further work can also be found in Section 4.8.

4.2 Evaluation of Results

In order to evaluate the performance of different techniques, the desired control points were manually positioned on a randomly selected set of 28 test images. Each technique was then used to process the images and the resulting points were compared to the manually positioned ones. The median error between manually and automatically positioned points was used as a metric to evaluate the accuracy of the search. The median error was used as opposed to the mean to reduce the effects of single points with very high errors on the overall score of a technique. Depending on how a given technique handles error conditions, erroneous points could be positioned as much as two meters away from their correct location. In this application, a point two meters away from the target is equally useless as a point ten centimeters away from the target. The median error value is less sensitive to very large error values for a small number of points than the mean error, so was expected provide a better metric for the usefulness of a give technique.

In order to provide a gold standard with which to evaluate the performance of the algorithms, the required points were manually positioned by an expert on the full resolution images. In the existing solution (described in Section 2.2), points are manually marked on a sub-sampled version of the image. Although no tests were run to evaluate the accuracy of these points, they are expected to have an maximum error of $2.5mm$ when compared to the gold standard, due to the reduction in resolution. All comparisons between the results of the algorithms and the existing solution are based on this expected maximum error.

4.3 Rule-based Method

Due to the fact that each step taken during the rule-based search relies of the correct completion of the step before it, the propagation of errors though the chain of rules was a significant factor, decreasing the overall accuracy of points found later on in the algorithm. The point for the top

of the head, which is the first point found, was the most reliably positioned of all, resulting in an median error of 24 pixels (5.7mm). In comparison, the points at the patient's wrists, being some of the last points found, and thus most susceptible to propagated errors had a much higher median error of 314 pixels (78mm). Overall, a median error of 152 pixels (38mm) was obtained across the 19 required points located using this method.

Another problem in using a rule based system is the large number of unexpected features present in the x-ray images. The most commonly occurring of these include oxygen bottles, stretchers and head restraints. Encountering these features breaks the assumptions made when designing the rule based system and frequently results in incorrectly positioned points. For example, if an oxygen bottle is placed across the patient's legs during a scan, the convergence between the bottle and the patients legs could be incorrectly identified as the patient's groin. This would result, not only in an incorrect groin position, but also the edge tracing algorithms for the bones in the legs would be initialized in the wrong place, greatly reducing the likelihood of correct matches for the points on the legs. Although it would be possible to design the system to account for one or two of these situations, the huge variety of possible arrangements of these non-patient objects makes this approach infeasible.

Because the rule-based method starts out with no prior model of the patient's shape, incorrectly positioned points are frequently placed in unlikely positions when considering the expected shape of the patient. As a result, this method is suited to having incorrect points detected though the use of the shape-based error detection method discussed in Section 3.6. Detecting these errors would allow the operator to be alerted to the error, and position the offending point manually. However, every image in the unseen test set had at least one point positioned incorrectly by the rule-based system beyond the 300 pixel threshold required by the error detection algorithm (see derivation later in Section 4.7). This would mean that in a practical application, the operator would frequently have to be correcting the errors of the program, which would reduce the time saving benefits of the system.

The other option when dealing with points detected as incorrect would be to perform additional processing in an attempt to correct the error. This processing could be done by re-running the offending section of the rule-based method using slightly different parameters for threshold values and edge tracing initialization. Alternatively, a completely different technique (possibly one of the other three discussed in Chapter 3) could be used to estimate the position of the incorrect point.

A further disadvantage of the rule-based system is its lack of flexibility: if any additional points need to be located, a new chain of rules will need to be constructed in order to locate them. For the same reason, it would be difficult to adapt this system to other applications. Nevertheless, a possible application of the bone tracing algorithm may lie in the automatic detection of fractures by searching for discontinuities or unusually high curvature of the bone's edge.

4.4 Rigid Template Matching

The rigid template matching was applied to all of the 27 unseen test images and proved to be the least effective of all tested techniques. The majority of points were positioned very far away from their manually position counterparts, resulting in a total median error of 610 pixels (126mm). This was likely because of the large variation in scale, texture and orientation of equivalent points across different images. This variation resulted in very blurred templates and therefore weak matches between the template and a new image. The points which performed best under this approach were those with high contrast transitions between body and background, such

as the top of the head and the shoulders. As a result, the median error for these points was significantly lower than average, 228 pixels in the case of the head, and 260 pixels for the shoulders.

Areas defined by more subtle changes in intensity, as in the case of the humoral head (Figure 3.5(b)) were more susceptible to the blurring effects caused by variations in the training set, and so produced less reliable matches (815 pixels median error for both humoral heads).

Due to this low accuracy, the rigid template matching search is not likely to be useful on its own for locating points on the patient. The performance of this technique could possibly be improved by adding a weighting factor to the match criteria based on the expected position of the points, however, as this method produced significantly worse results than the others, so the improvements from any modifications would have to be considerable in order to make this method worth while. In the end, this system doesn't adequately account for the large variations present in the x-ray images so is unlikely to perform well without serious modifications.

4.5 Active Appearance Models

When initialized using the basic registration, the active appearance model did not perform as well as the rule based method, producing a median error of 187 pixels (46mm) per point compared to 152 pixels (38mm) for the rule based system. However, if the model was initialized based on the results of the rule-based search, the accuracy would be improved to 88 pixels (21mm) per point.

The active appearance model search was tested with a variety of different combinations of model complexity, resolution, tessellation and registration techniques. The most effective combination of parameters was found to be:

- 92 point model.
- 256^2 pixels model resolution.
- Manual tessellation with background triangles preserved.
- Model registration based on the results of the rule based search.

Using these parameters, the search resulted in a median error of 88 pixels (21mm) per point. The effects of changing these parameters are outlined in the following sections.

One of the advantages of this system is its flexibility: if any additional points need to be found, the only change that needs to be made to the system is the labelling of this point in the training set. This method could also be applied very easily to finding the configuration of any objects in the image, provided a set of annotated training images and an adequate registration technique are available.

One parameter not considered when testing the AAM search, was the magnitude of the parameter offsets used for linear regression when building the Jacobian matrix (Section 3.4.3). As this value would very likely effect the accuracy of the final position of the model, it may be worth further investigation.

Stegmann [7] implemented a second stage optimisation which was run once the model had converged. This method was found to improve the accuracy of the final match and it is expected that similar improvements would be possible in this application.

The procedure for comparing model configurations that led to the definition of the above model as the best one, is described in the following sections.

4.5.1 Performance of Different Models

A number of different model configurations (as described in Section 4.5) were attempted, based on different combinations of

- 23 point model and 92 point model;
- Manually suppressing non-patient features in the training images or using the original image;
- Delaunay or manual tessellation;
- Deleting or preserving background triangles.

Figure 4.2 shows the mean error for all points for a number of combinations of these configurations.

The combination resulting in the least error, was a detailed model built from the original images, manually tessellated, preserving all triangles. Attempts to suppress background objects, deleting the background triangles and erasing background features from the training images, resulted in poorer performance, especially when searching in images with commonly occurring background objects such as the head restraints shown in Figure 3.7.

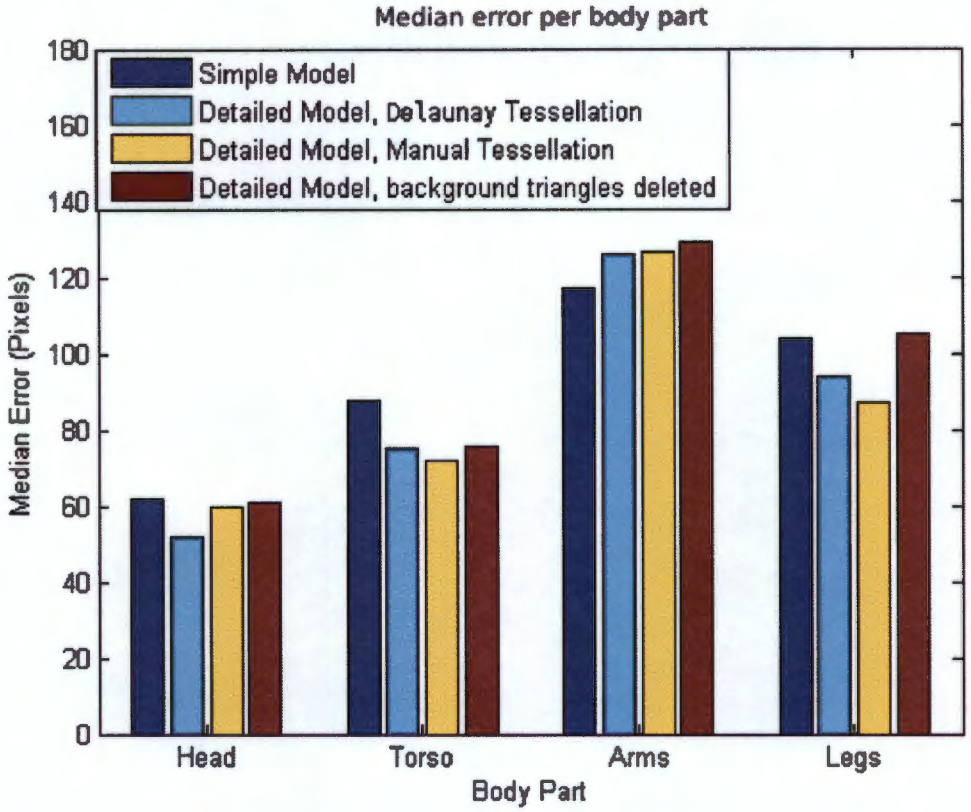


Figure 4.1: Effect of model on search accuracy.

4.5.2 Effects of Different Registration Techniques

In order to evaluate the effects of different registration techniques, each technique was used to initialize the AAM search and the accuracy of the search results were used to evaluate the effectiveness of the registration. The results are shown in Figure 4.2.

Unsurprisingly, the best results were obtained from registering the model based on operator-annotated points. The small errors present in this case are due to constraining the operator positioned points to shapes found in the training set, and to non-linearities in the relationship between the error-function and point-error-distance at very small offsets from the ideal fit, causing the model to converge on a sub-optimal solution.

The basic registration using the patient's size and center of mass, while resulting in comparatively poor performance for points on the limbs, produced more accurate searches than all other registration techniques tested when considering points on the torso only (64 pixels median error compared to 81 pixels for registering using the rule based points and 69 pixels using the operator positioned points). It is interesting to note that basic registration is more accurate registration based on manually positioned points. This is likely to be due to the variation in the initial position of the limbs. When the initial registration of the limbs is incorrect (as would happen with the basic registration), changes to the shape and texture of limbs would have less effect on the overall error function, thus allowing the sections of the torso greater flexibility in matching the search image. Based on this observation, the same search was performed using a model with all points on the limbs deleted; this model produced similar improvements in quality independent of the registration technique used.

When comparing the model fit, before and after running the search on an inaccurately initialized model, it is interesting to note that running the search did not improve the results of model fit. This illustrates the active appearance model's propensity for converging on an incorrect solution if the initial registration is not good enough.

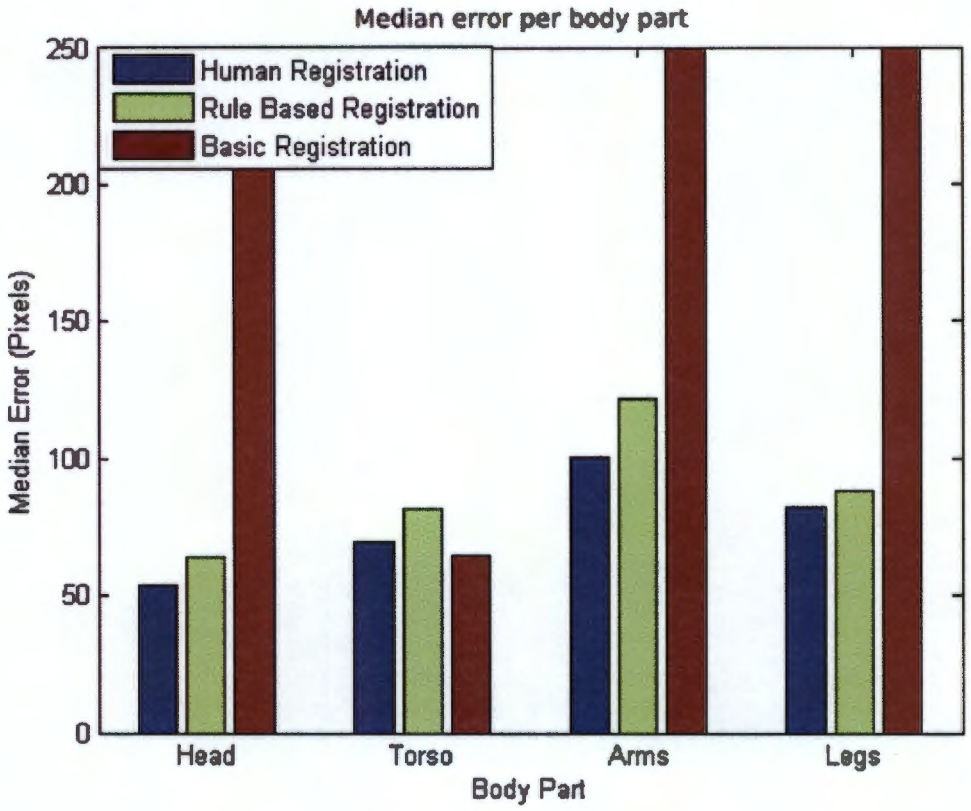


Figure 4.2: Effect of registration technique on search accuracy.

4.5.3 Effects of Model Resolution

The effect of building the appearance model at different resolutions was considered by running the search with different resolution models and comparing the accuracy of the results. Due to limitations in the image warping implementation, only resolutions of 2^{2n} up to 512^2 could be used. Figure 4.3 shows the effects of changing model resolution on the resulting median error across all tested images.

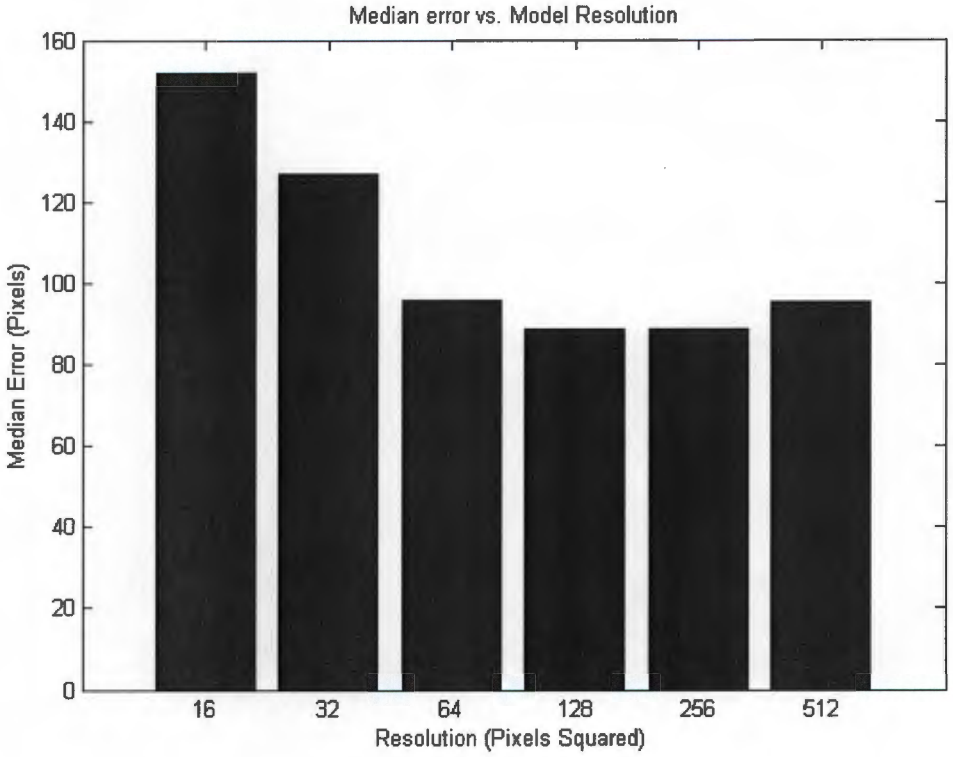


Figure 4.3: Effect of model resolution on search accuracy

Cootes et al [18] implemented a multi-resolution AAM search which would search in progressively higher resolution versions of the image. This was found to improve the speed of execution of the algorithm as well as reducing the chance of the model diverging when the initial registration is not sufficiently accurate. When searching for the patient, none of the training images caused the model to diverge when initialized from the rule based method, so a multi-resolution search may not add any value in terms of accuracy.

4.6 Articulated Active Appearance Models

As the first stage of the articulated active appearance model (AAAM) search (a model of the torso only) is equivalent to optimizing a standard active appearance model of the torso only, it is not surprising that the accuracy of points on the torso demonstrated similar improvement in accuracy over the full body AAM model (64 pixels median error compared to 84 pixels) as previously experienced when using an AAM model with the limbs deleted (Section 4.5.2). However, the models representing the limbs would frequently diverge during the full-body AAAM search, resulting in a lower overall accuracy for these points when compared to the standard AAM search.

The poor performance of the limbs is surprising given the reduced amount of variation needed to be expressed by the model when compared to a full-body AAM search. This may be in part due to the propagation of errors though the model hierarchy (as previously experienced during the rule-based search, Section 4.3). In the case that a parent model results in an inaccurate match, the registration of the child models may not be good enough to ensure convergence. Figure 4.4 shows the average median error for all points depending on their position in the hierarchy, it is clear from this figure that points lower down the hierarchy are more likely to be inaccurately positioned, which is to be expected if errors from the parent models are effecting the accuracy of their children. An option to solve this problem would be use an additional registration step for the models representing the limb segments. The bone tracing method described in Section 3.2.3 may prove be of use in this regard. A second option may be to use a multi-resolution search to reduce the probability of the models diverging due to a poor initial registration.

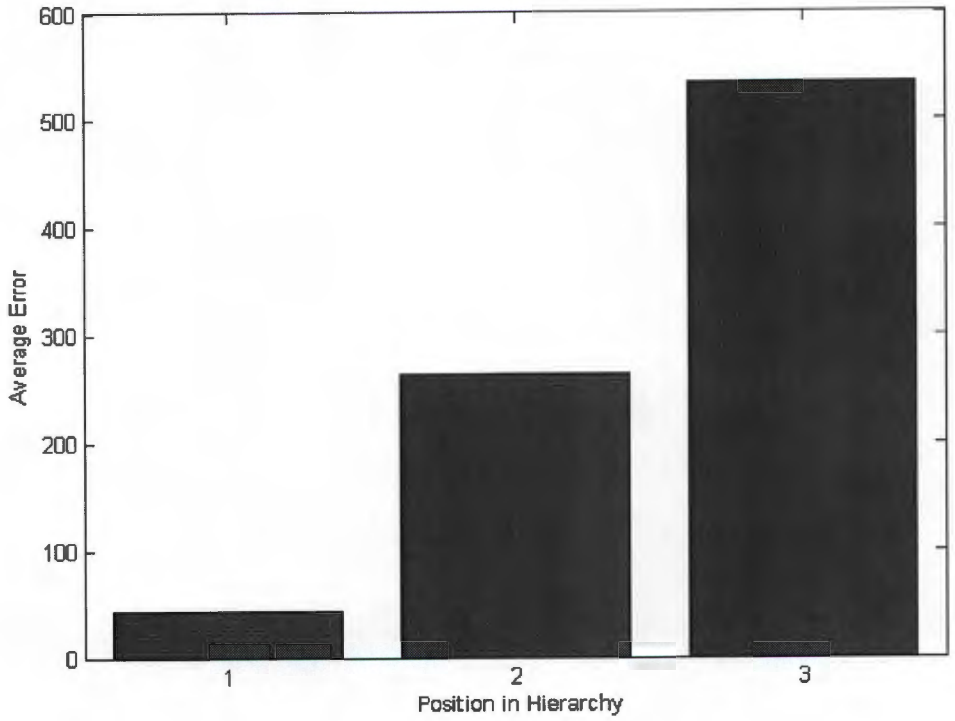


Figure 4.4: Point accuracy vs. position in the search hierarchy, where Position 1 includes all points on the torso, Position 2, all points on the head and upper limb segments and Position 3, points on the lower limb segments.

When initialised from operator positioned points, the models for the limb segments still produced less accurate results than equivalent AAM (initialised from operator positioned points, see Section 4.5.2). Considering this, it is likely that the propagation of errors though the hierarchy is not the only cause of the poor performance of the limb segments in this search. Another possible cause may be that the 0.5 standard deviation parameter offset used to build the Jacobian matrix used in the model optimisation stage (Section 3.4.3), is not suitable for the limb segment models.

4.7 Incorrect Point Detection

4.7.1 By Shape

To evaluate the feasibility of detecting errors using the expected shape, a single point from the human annotated point set was given a random offset. The shape-based incorrect point detection algorithm (Section 3.6) was then applied to the point set to produce an error metric for every point in the set. Figure 4.5 shows the resulting error metric from 1000 test cases with induced errors varying from 0 to 800 pixels. In the figure, the red crosses represent the error metric associated with the incorrect point, while the blue circles represent the maximum error associated with the correctly positioned points. As shown by this figure, it appears be possible to separate the incorrect points from the correct ones based on this error metric for all errors greater than 300 pixels.

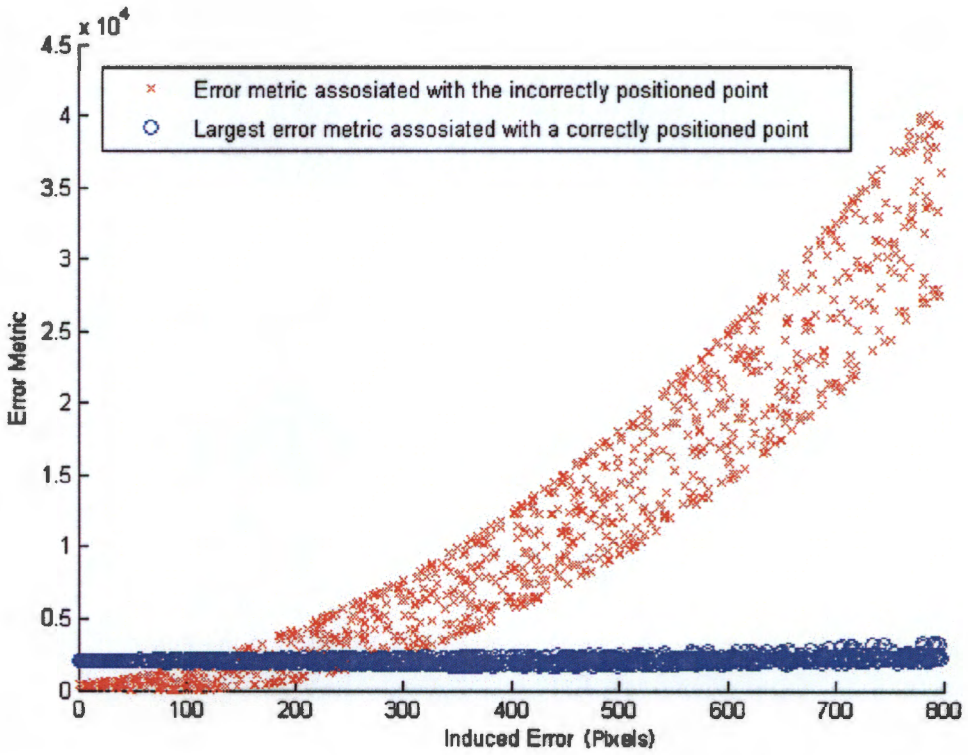


Figure 4.5: Shape based error metric vs. the magnitude of induced error in pixels.

To evaluate the effect of having multiple incorrect points in the data set, a random error of 400 pixels was introduced into a random number of points. The results of the error detection algorithm were recorded from 1000 different test cases. Figure 4.6 shows the percentage of correctly labelled points vs the number of incorrect points in the data set.

Finally, random errors were introduced into a random number of points; the probability of correct labelling of the erroneous point vs. the magnitude of the error is shown in Figure 4.7. As shown in the graph, the accuracy of identifying incorrect points is dependent on the amount of error between the point and its correct position. This is to be expected as the farther the point is away from the correct position, the greater the change induced in the shape vector. The greater the change to the shape vector, the more likely it is to be beyond the threshold distance from the mean shape.

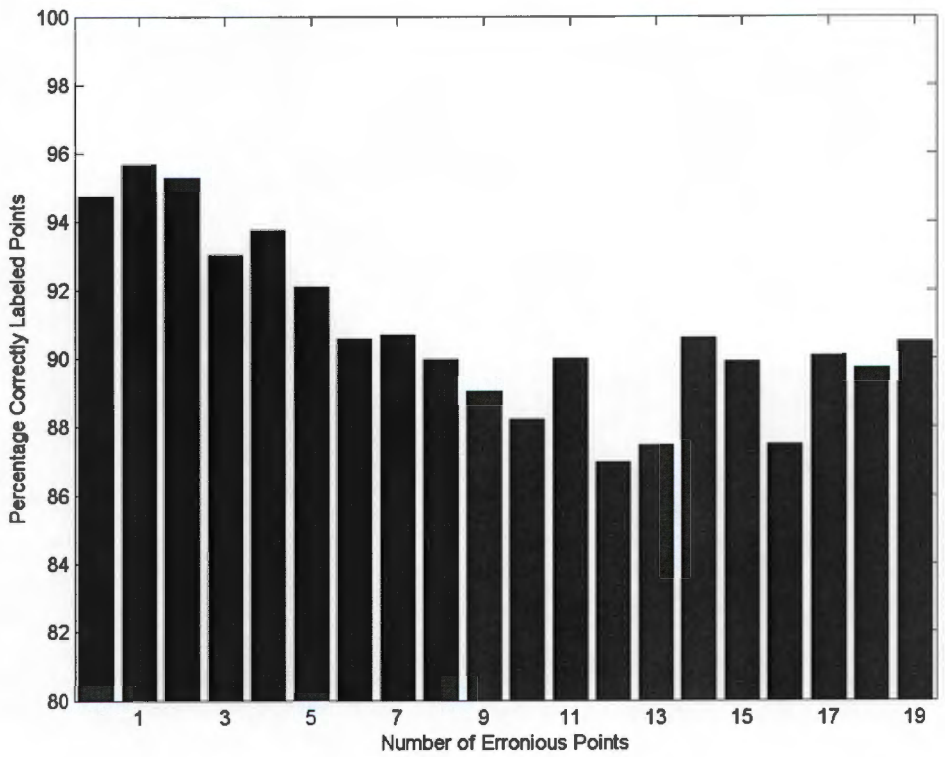


Figure 4.6: Percentage of correctly labeled points vs. number of incorrect points.

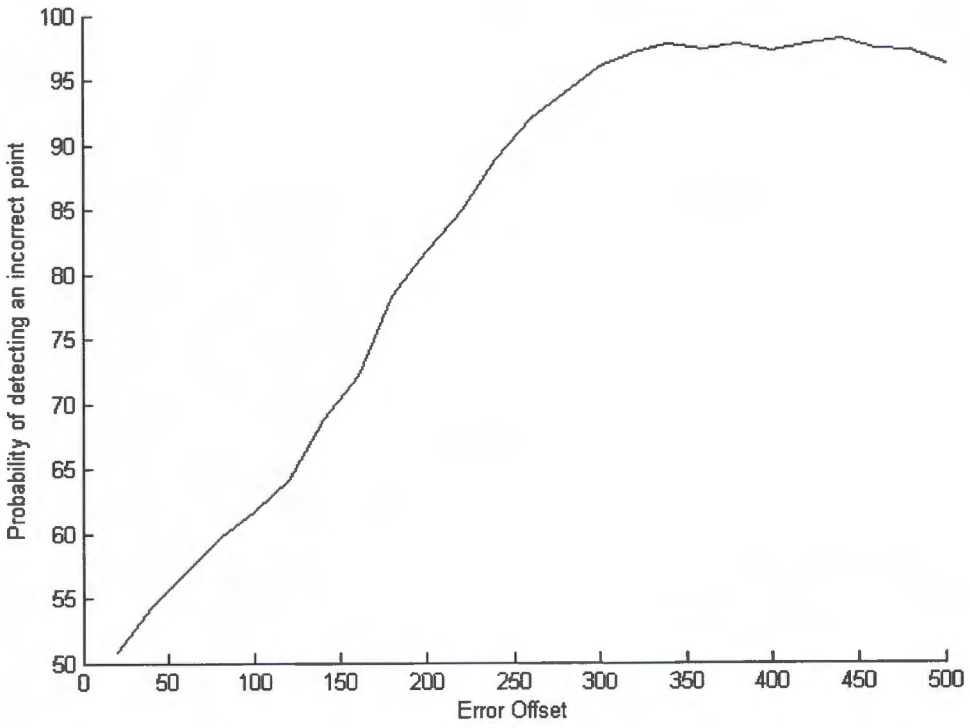


Figure 4.7: Probability of detecting an incorrect point using shape based error detection vs induced error magnitude.

Error detection by shape is not possible for results generated by the active appearance model search (Section 3.4) as the points produced by this search are already constrained by the shapes found in the training set.

4.7.2 By Texture

The texture based error detection algorithm was tested in a similar fashion to the shape-based algorithm. A random error was added to a single point in a correctly labelled set. The texture based error detection algorithm (Section 3.6) was then used to generate an error metric for each point in the set. The error metric generated for the single incorrect point was then compared to the maximum error metric generated for all the correctly positioned points.

The results of 100 point sets tested in this fashion is shown in Figure 4.8. In the figure, the red crosses represent the error metric associated with the incorrect point, while the blue circles represent the maximum error associated with the correctly positioned points. Unfortunately, in the majority of cases, the error metric associated with the incorrect point was no greater than that of the correct points. This method of error detection was judged not to be useful as no threshold could be found to separate the incorrect points form the correct ones based on the greyscale error metric.

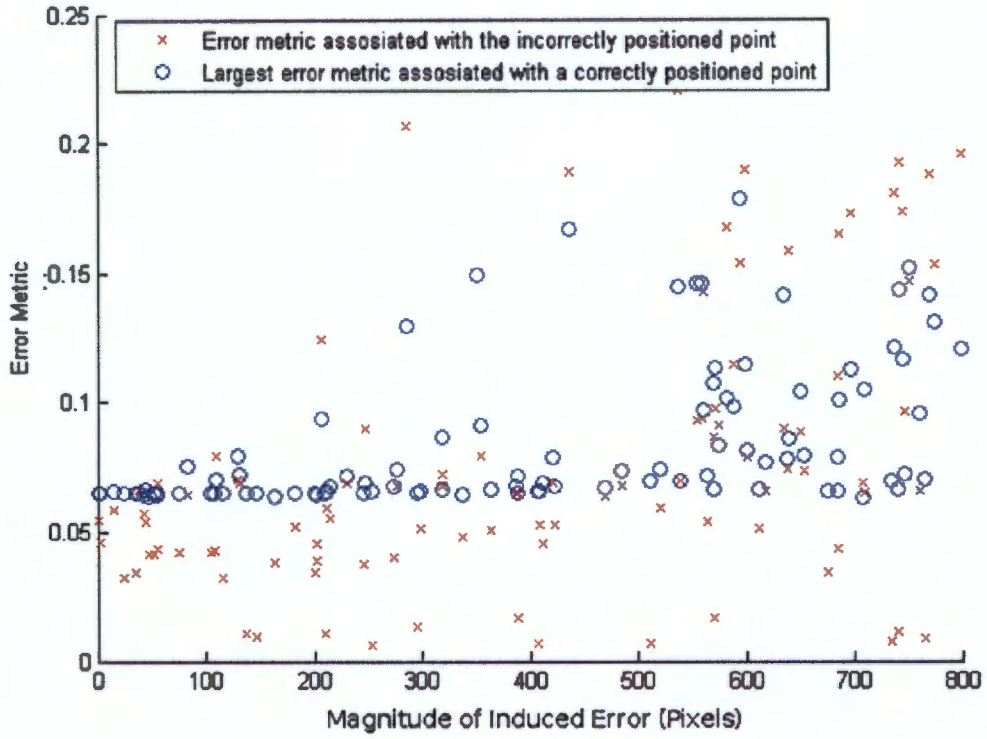


Figure 4.8: Texture error metric for incorrect points (circles) and correct points (crosses) vs. magnitude of induced error

4.8 Conclusions

Figure 4.9 shows a comparison of the median error of each search method for the different parts of the body. The AAM search presents the lowest overall error of a median 88 pixels (21mm) per point positioned, so this method is recommended for finding points for the application of splitting the patient into smaller regions. For the purpose of segmenting the full-body image into smaller regions of interest, it should be tested whether this error is small enough to produce useful sub-images. The redundancy of points in any desired region may allow adequate segmentation in spite of the error.

For the application of selecting control points for a lateral spine scan using automatic technique factor control (ATFC), only a subset of the total points is required. In addition, only the error component in the direction of motion of the C-arm need be considered. Taking these restrictions into consideration, the most accurate search method is the articulated active appearance models, which produced a median error of 24 pixels (5.7 mm) across all the relevant points, in the relevant direction. All of the points required for ATFC are contained in the head and torso model so the models representing the limbs need not be considered in a practical application. As there is very little rotation of the head with respect to the torso present in the test images, it is likely that a standard AAM model of head and torso alone, would perform equally well in this application, however this was not tested.

When compared to the current system requiring the operator to manually mark the control points for a lateral spine scan, only 20% of the points positioned by the articulated active appearance model search were as accurate or more accurate than the theoretical accuracy of the operator positioned points in the sub-sampled image (Section ??)). It is unlikely, however, that

the human operator will provide pixel-perfect positions for the control points. An investigation into the average accuracy of the human positioned points using the existing method for ATFC would be needed in order to properly compare the two methods.

As the first stage of the active appearance model produced the most accurate matches for points on the torso, 64 pixel (16mm) median error per point, this method was used for positioning the algorithms considered in Part 2.

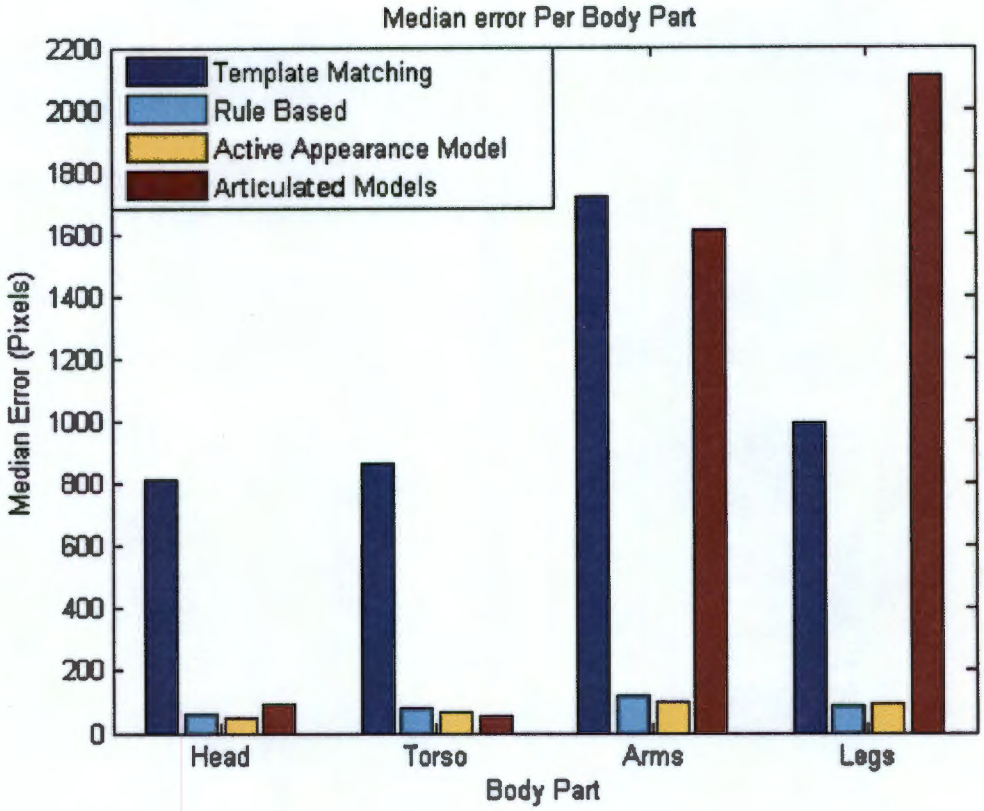


Figure 4.9: Overall Errors for all search techniques.

As shown in Figure 4.9, the different segmentation techniques result in different accuracies for different parts of the body. It may be possible to use a combination of techniques and select the results based on which method is most likely to give an accurate result for the given point. However, the effectiveness of this technique would have to be evaluated on an separate data set to ensure that the relative performances of each technique are not unique to this data set.

An important factor for a practical application which was not considered in this study, is the speed at which the search algorithms converge. This should be investigated before a system is implemented.

The poor accuracy demonstrated by all techniques in finding the arms may be due in part to the reduced contrast in these areas relative to the rest of the body. It may be possible to improve results by applying a filter to normalize the contrast before running the search. One such filter is the local normalisation filter (described in more detail in Section 6.2.2) which normalizes local contrast by subtracting from every pixel the average greyscale intensity of all pixels in its local vicinity then dividing the result by the standard deviation of those pixels. This would allow errors on the arms to contribute a greater value to the overall error metric in the AAM based searches. The same filter would need to be applied to the training images

before the model is built.

Part II

Airway Segmentation

Chapter 5

Background

Due to new drug resistant strains and cross infection with HIV, the number of people dying from tuberculosis is increasing. In South Africa, death rates as a result of tuberculosis increased from 89 deaths per 100,000 population in 1990 to 134 deaths per 100,000 population in 2004 [19].

The gold standard for the diagnosis of pulmonary tuberculosis (PTB), is a cultured sputum sample examined microscopically for the tuberculosis bacteria (*M. Tuberculosis bacilli*). However, this process can take up to four weeks to complete and it is important that tuberculosis cases be identified early and isolated to limit the spread of infection [20]. In addition, sputum samples from children are difficult to obtain and are less likely to present bacilli [21]. Abnormalities in a standard chest radiograph can be suggestive of PTB infection but the features presented can vary greatly between cases and often masquerade as other maladies [20, 22]. Tuberculosis is known to affect the bronchi in the following ways:

- Bronchiectasis (dilation of the bronchi) can be brought on by the scarring of the bronchial walls [23].
- Necrotic material from infected nodes can be discharged into the lumen [22].
- Infected lymphatic nodes can compress the bronchi [22].
- It is probable that the pressure exerted on the bronchi by enlarged lymph nodes will alter the shape of the airways beyond simply compressing them.

Despite this, the shape of the pulmonary airways is not routinely considered when chest x-rays are examined for PTB, possibly due to the poor visibility of the airways on traditional x-rays. Recent advances in radiographic technology, implemented in the Lodox StatScan machine, allow for improved visibility of airways [2] raising the possibility of considering airway morphology in the diagnosis of PTB.

5.1 Problem Description

Before any analysis of the airways can be performed, they first need to be located in the image. This is still a challenging and time-consuming task, despite the improved contrast found in the Statscan radiographs.

The pulmonary airways can be identified visually on a chest radiograph as a tree-like structure with an interior slightly darker than the surrounding image (Figure 5.1 shows an example of the chest region of a full body scan produced by the Lodox Statscan machine, including the pulmonary airways). Although the airways are relatively clear over the denser regions in the chest such as the heart and spine, over the darker regions the edges of the airways are generally

badly defined and in places unidentifiable by the human eye. It would be useful to at least partially automate the process of airway segmentation to improve the speed and accuracy of human identification of the airways.

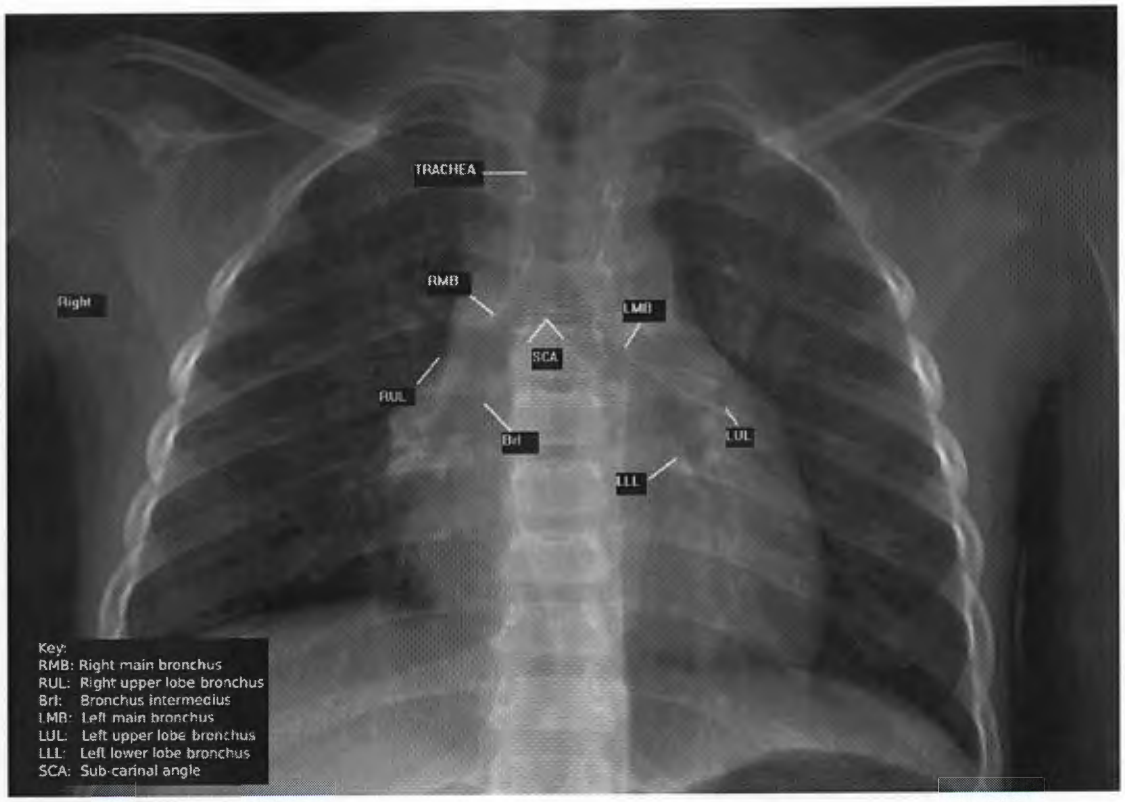


Figure 5.1: Chest region of a Lodox full body scan.

Figure 5.2 shows the histogram of pixel intensities, for bone, airway and lung tissue, in the chest region, averaged over 10 images. Although the airways are darker than the local surrounding areas, the large amount of overlap between the intensity of the classes prevents the use of simple thresholding filters. A successful segmentation technique will need to consider more information than the greyscale intensity of the pixels in order to remove the ambiguity caused by this overlap between classes.

Another possible parameter which could be considered by a segmentation algorithm is the shape of the airways. Figure 5.3 shows a selection of airway shapes manually extracted from x-ray images of different patients. As shown in the figure, the airways do present a similar shape but the amount of the airway tree visible, even to a human operator, is highly variable resulting in a number of different configurations.

5.2 Previous Work

Andronikou et al. [25] performed a study of chest CT scans of children with pulmonary tuberculosis. Of the cases examined, 46% presented lymphadenopathy (enlarged lymph nodes) which resulted in bronchial compression in 35% of the total cases. No quantitative measure of bronchial compression was used, rather the presence of compression was assessed subjectively by a radiologist. It is possible that more accurate, repeatable measurements would reveal a larger proportion of lymphadenopathy cases presenting bronchial compression.

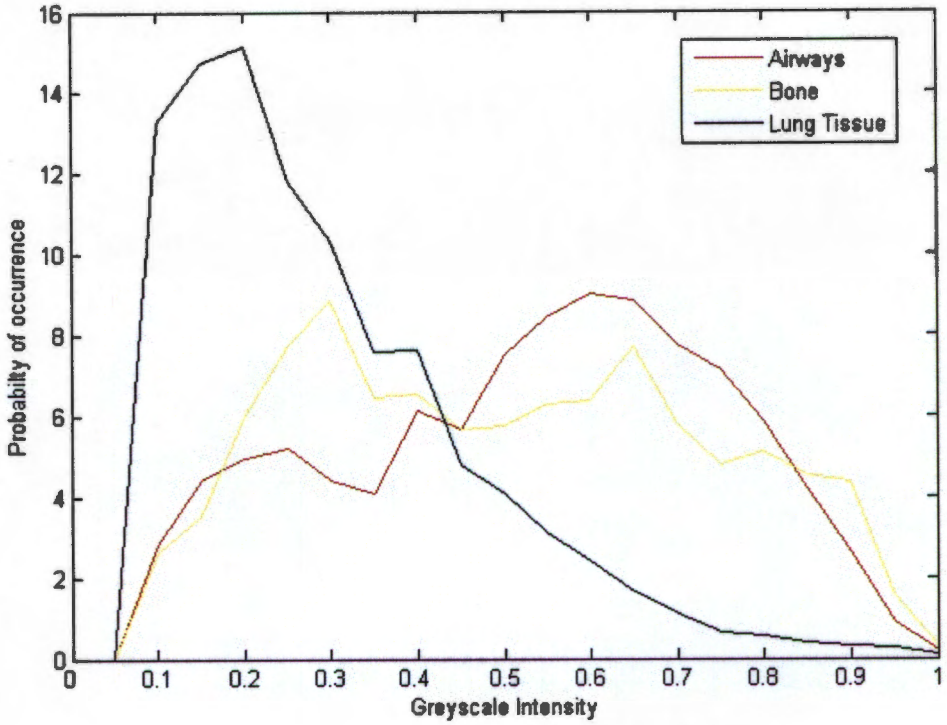


Figure 5.2: Histogram of chest region greyscale intensity.

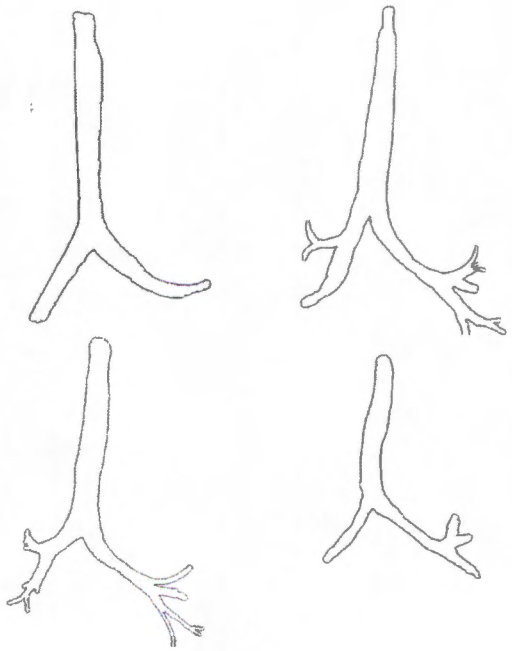


Figure 5.3: Manually segmented airway shapes.

In a similar study involving adults with acute PTB [23], evidence of lymphadenopathy was found in only 11% of of the sample group, 17% having brochiectasis. Unfortunately bronchial compression was not considered.

While Andronikou et al. [25] found that detecting the enlarged lymph nodes themselves was more reliable than detecting the nodes' effects on the airways from a CT scan, the enlarged nodes are not visible on anterior-posterior x-ray because of the reduced contrast between soft tissue types on x-ray compared to CT.

In addition to the reduced contrast between soft tissue types, using an AP x-ray has several other disadvantages over CT. Specifically the high resolution and three dimensional nature of CT allows for visualization of structures that may not be visible though plain radiography. However, these disadvantages are compensated for by the significantly lower radiation dose required for a Statscan radiograph. The Statscan machine typically delivers an entrance dose of 0.05mSv per paediatric chest image [24] compared to the 6.5mSv required for a CT scan of a similar region [26]. This is of particular significance when dealing with children as the probability of them developing radiation induced cancer is greater than that of adults [27].

5.2.1 Computer Assisted Methods

Very little work has been done on computer assisted detection of tuberculosis signs in chest radiographs. Sarkar and Chaudhuri [28] attempted to measure the progression of tuberculosis through image processing techniques applied to digital chest x-rays. Areas of cavitation and infiltration were segmented, then the total area of these sections was compared across a series of radiographs to quantify the progression of the disease. Neither airway morphology nor lymphandenopathy were considered.

Koeslag [9] used template matching to detect the lesions associated with millary TB, but did not consider the airways.

5.2.2 Segmentation techniques

To the author's knowledge, all previous studies on segmentation of airways have been conducted on CT images, as in the work of Tschirren et al. [29], in which the cross sections of the airways were segmented in a lateral CT slice, after which the segmented cross sections were stacked together to form a three dimensional model. This technique is obviously not applicable to standard x-rays as the third dimension is unavailable, so it is not possible to obtain a cross-sectional view of the airways.

Although the segmentation of the airways has not been considered in anterior-posterior chest x-ray images, a variety of techniques have been used successfully to segment other features in chest radiographs.

Kass et al. [11] introduced what they called "snakes", curves which are fitted to features in an image to minimize an energy function. After initialization, the curve is iteratively improved by moving control points and re-evaluating the energy function. The energy function usually contains a measure of the curvature of the snake and the strength of image edges beneath it so that the minimum value occurs when the snake fits to the strongest edges while maintaining a smooth curve. Snakes may prove to be useful in this application as the edges of the airways do appear to present a smooth curve transitioning between the darker interior of the airway and the rest of the image. However, the edges of the airways are by no means the most intense edges in the image; the borders of ribs, spine and heart all present much stronger edges than the airway boundary; this could confuse a naive snake implementation.

Active shape models, introduced by Cootes et al. [13], allowed a snake-like model to be constrained based on its shape, which prevents the model from fitting to any implausible shapes. These have been used in [14] to detect the lung fields in chest posterior-anterior chest radiographs. When applied to the problem of airway segmentation, constraining the model to fit known shapes may prevent the snake from fitting to the stronger edges of ribs and spine; however, this technique requires that the shape of the object being searched for is known and fixed. In the case of the airways, large sections forming part of the airway shape may not be visible, even to the trained expert (as shown in Figure 5.3). This could possibly be overcome by only considering the subset of the airway volume which is visible in every image across the training set but a brief inspection of the segmented training set (Figure 5.3) reveals that this subset would likely only include the trachea and first two branches, so may be of limited diagnostic value.

Active appearance models, previously described in Chapter 2.5.6, were considered for segmentation of the airways. However, to build an accurate model of the greyscale texture, similar textures are required between each instance of the shape. The greyscale values of the airways are largely defined by the anatomical features behind them, namely ribs, spine etc. As the patient breathes, the lungs and airways will move relative to these fixed anatomical features changing the texture of the airways in the images. This movement will limit the effectiveness of texture based techniques; however appearance models could still be useful in locating the patient in the image, and providing an initial position for the airway model.

Expectation maximisation segmentation [30], is an iterative segmentation technique which seeks to assign pixels to classes by finding the maximum likelihood for a set of parameters which control the classes. The parameters for each class are the mean and standard deviation of a normal distribution, which is used to approximate the probability distribution of greyscale values belonging to the class. In the final segmentation each pixel is assigned to the most likely class based on the probability function for its greyscale value. Moon et al. [30] extended this method to include prior knowledge in the form of a probability map of expected class vs. pixel position. The probability function used in the expectation maximisation functions were extended to include this probability map in addition to the greyscale values of each pixel.

Schilham et al. [31] used local normalisation as a pre-processing step to remove the ribs to improve their lung cancer nodule segmentation algorithm. This filter normalised the contrast throughout the image, reducing the effects of low frequency variations in image intensity (such as those effected by spine and heart shadows).

Chapter 6

Methods

6.1 Chapter Summary

This chapter details the methods used to segment the airway tree from the chest region of the x-ray image. Images are first normalized with respect to the shape of the lung fields (Section 6.2.1) in an attempt to reduce the amount of variation present in the shape of the airways. The images are then filtered (Section 6.2.2) to improve the visibility of the airways. Finally two segmentation techniques are applied to the filtered, shape normalized images, expectation maximization classification (Section 6.3) and locating the airway edges using snakes (Section 6.4).

6.2 Pre-Processing

This section presents the pre-processing methods which were used in an attempt to improve the visibility of the airways before applying the segmentation algorithm. Two methods were applied at this stage:

- Shape normalisation, which attempts to remove any variation in the airways dependant on other features in the image.
- Local normalisation, which attempts to increase the contrast of the airways as they transition from the high intensity background over the spine and heart onto the lower intensity regions in the lung fields.

6.2.1 Shape Normalisation

A large amount of the variation of airway shape, size and position present in the training set can be accounted for by the shape and position of the rest of the patient. If some of this variation can be removed by considering the shape and position of features around the airways, the segmentation problem will be much easier to solve. However, it is important that any warping of the image done at this stage is reversible so that the results of the segmentation algorithm can be transformed back onto the original image.

In an attempt to remove some of this variation, the images of the lung fields were first projected onto a shape and position normalised space. To accomplish this, an active appearance model of the patient's torso (described in Section 3.4) was used to locate the lung fields in an image. The image of the lungs could then be warped onto the mean lung shape to reduce the amount of variation in the dataset not due to the airways themselves. A more detailed model

of the torso was built compared to the one used for locating the patient in Section 3.4, (for a full list of landmarks used in this model, refer to the Appendix) so that more information could be included about the edges of the lung fields. An image of the torso annotated with the control points used by the model can be seen in Figure 6.1. The position of the first bifurcation point, at the lower extremity of the trachea (marked in green in figure 6.1) was included in the model in addition to the points used previously.

Figure 6.1: Landmark points used to locate the lung fields.

When presented with a new image, the appearance model was first optimised to fit the image, using the methods described in Section 3.4.3. Once the search algorithm was complete, the image was warped to align with the lung field shapes found in the training images. The following different image warping techniques were considered:

- Simple alignment
- Affine warping
- Piecewise linear warping based on the boundary of the lung fields
- Piecewise linear warping including the first bifurcation point

"Simple alignment" only removes the effects of rotation and scaling by finding the best alignment between training shapes. "Affine warping" warps the image using an affine transform. "Piecewise linear warping" warps the lung fields to align every point identified on the respective lung fields. "Piecewise linear warping including the first bifurcation point" uses the same warping technique as the piecewise linear warping, however the location of the first bifurcation point (as estimated by the active appearance model search) is included as a control point during the warping.

All these techniques were tested using both manually annotated points and points found using the AAM search (see Section 3.4). Each image in the training set was aligned using

each technique and the resulting warped images were used to generate a probability map of the airway position, for each warping technique. The probability of a given pixel in the shape normalised space was estimated using the number of occurrences in the training set which resulted in the given pixel being part of the airway. The probability map of airway position is thus calculated as a simple average of the manually segmented, shape normalised airway images:

$$P_{airway}(x,y) = \sum_{i=1}^{n} A_i(x,y)$$

where $P_{airway}(x,y)$ is the probability map, n is the number of images in the training set and $A_i(x,y)$ is a manual binary segmentation of training image i into airway and background classes.

An effective shape normalization technique will remove the largest amount of variation from the training set before the segmentation, thus an effective technique should align the manually segmented image more effectively resulting in a sharper probability map. In an attempt to evaluate the effectiveness of these warping techniques, before applying any segmentation algorithm, the sharpness of the resulting probability map was used as a measure of how well the shapes in the training set were aligned using the given warping technique. The image sharpness metric used was the "Energy of the image Laplacian" [32], calculated as:

$$s = \sum_{i,j}(4x(i,j) - x(i-1,j) - x(i+1,j) - x(i,j-1) - x(i,j+1))^2$$

where $x(i,j)$ is the pixel at position (i,j) in the image in question, and s is the resulting sharpness metric. In this fashion, the metric represents the sum of the difference between each pixel and its immediate neighbours, thus giving higher values for images with more high frequency variation.

6.2.2 Filtering

As evident in Figure 5.1, a large amount of the image intensity variation present around the airways is due to unrelated structures such as the spine and heart shadows. In an attempt to reduce the effects of these unrelated structures on the greyscale intensity, a local normalization filter was applied as described by Schilham et al. [31].

The filtered image, $g(x,y)$ is calculated as:

$$g(x,y) = \frac{f(x,y) - \mu_f(x,y)}{\sigma_f(x,y)}$$

where:

- $f(x,y)$ is the input image.
- $\mu_f(x,y)$ is the local image mean at point (x,y).
- $\sigma_f(x,y)$ is the local image standard deviation at point (x,y).

The local mean, $\mu_f(x,y)$, and standard deviation, $\sigma_f(x,y)$, can be estimated as follows:

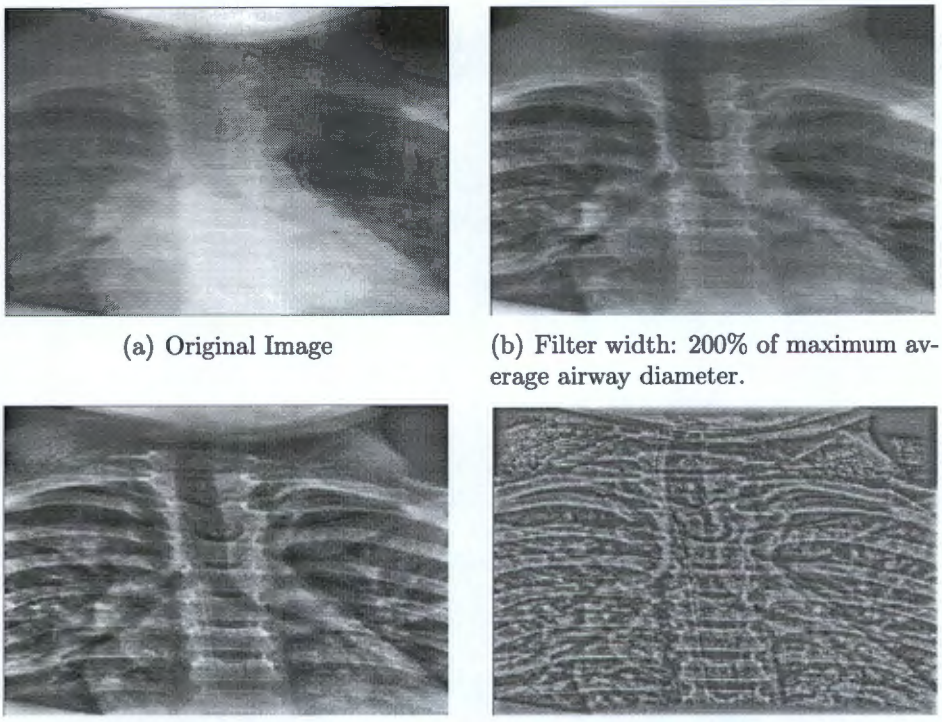
$$\mu_f(x,y) = G(f(x,y))$$

$$\sigma_f(x,y) = \sqrt{G(f(x,y) - \mu_f(x,y))^2}$$

where $G(f(x,y))$ represents a Gaussian blur applied to the image function f.

The only parameter needed for this filter is the standard deviation of the Gaussian blur, which determines the size of the region used in calculating the local mean and deviation. The effects of changing this parameter on the resulting image can be seen in Figure 6.2.

As the filtering is performed in the shape normalised space (described in Section 3.4.2), and as each image can potentially undergo different amounts of scaling when being projected into this space, it is not possible to calculate the spacial resolution of the warped image. So, in order to provide a scale independent measure of distance, filter widths were expressed in terms of the "average maximum airway diameter", that is, the maximum width of the airways averaged over each image in the training set when projected into the shape normalised space.



(a) Original Image

(b) Filter width: 200% of maximum average airway diameter.

(c) Filter width: 100% of maximum average airway diameter.

(d) Filter width: 20% of maximum average airway diameter.

Figure 6.2: Effects of changing the filter width of the local normalising filter on the resulting image.

Parameter Optimisation

In order to find the ideal parameters to use for the filter, the filtered images were used to generate metrics describing the suitability of the image for classification. Two metrics were used, one determined by the relative greyscale intensities between the airways and the rest of the image, and the other based on the relative strength of the edges of the airway compared to other edges in the image.

To calculate an edge strength metric, the edge strength at each pixel was estimated using a Sobel edge filter [32]. The edge strength metric was calculated as the ratio between the average edge strength for all pixels within a two pixel distance from the edge of the manually annotated airway edges, to the average edge strength of all pixels in the image.

When considering segmenting the airways based on the darker greyscale intensity, the following metric, M, was used to evaluate the separability of the airways from the background:

$$M = \frac{\overline{(\mu_1 - \mu_2)}}{\sigma_1 \sigma_2}$$

where μ_1 and μ_2 are the mean intensity values for the set of pixels representing the airways and the background respectively and σ_1, σ_2 are the standard deviations present in the airway and background pixels.

6.3 Expectation Maximisation Classification

Expectation maximisation (EM) classification was used by Moon et al.[30] to segment tumours from healthy tissue in brain CT images. This algorithm seeks to remove ambiguity in a greyscale threshold by considering the relative probability of each class occurring depending on the position in the image. This section describes this algorithm to the segmentation of the airway images after they have been shape-normalised.

A set of 20 training images were manually segmented into airway and non-airway regions, providing a binary image of pixels labelled as airway or non-airway. Each of these manual segmentations were aligned and warped onto a normalised shape using the methods described in Section 6.2.1. The segmentation algorithm as described by Moon et al. [30] is divided into classification and parameter estimation steps which are iterated to converge on the optimal solution.

In the classification stage, the probabilities of pixels belonging to each group us calculated based on the pixel intensity and a pre-calculated probability map of class position (details on the construction of a probability map for airway position can be found in Section 6.2.1). The probability map for each class of pixels is calculated as follows:

$$p(T_i = j | y_i, \theta_j) = \frac{p(y_i | T_i = j, \theta_j) p(T_i = j)}{\sum_k^n p(y_i | T_i = k, \theta_j) p(T_i = k)}$$

where:

- T_i is the class assigned to the pixel at position i
- y_i is the intensity of the pixel at position i
- θ_j are the parameters (μ_j and σ_j for class j)
- $p(T_i = j)$ is the probability of pixel i belonging to class j as estimated by the prior model and
- n is the number of classes of pixels in the image.

The probability distribution function of each class of pixels, $p(y_i | T_i = j, \theta_j)$, is modelled as a Gaussian distribution with mean and standard deviation parameters.

$$p(y_i | T_i = j, \theta_j) = G_{\sigma_j (y_i - \mu_j)}$$

The parameters for each class (μ_j and σ_j) are updated in the parameter estimation step as follows:

$$\mu_j = \frac{\sum_i y_i p T_i = j | y_i, \theta)}{\sum_{ip} (T_i = j | y_i, \theta)}$$

$$\sigma_j^2 = \frac{\sum_i p(T_i = j|y_i, \theta)(y_i - \mu_i)^2}{\sum_i p(T_i = j|y_i, \theta)}$$

Where $p(T_i = j|y_i, \theta)$ is the probability of a pixel at position i with greyscale intensity y_i belonging to class j as calculated in the classification step.

The classification and parameter update are repeated until a stable solution is reached. The image is then classified based on the class probabilities at each pixel.

Two different applications of this technique were considered. The first attempted to segment the image into two classes; airway and non-airway, but due to the large amount of variation in greyscale found on the airways, a second was considered which attempted to segment the image into 4 classes of descending greyscale intensity; bone and soft tissue, airway (over the heart and spine shadows), lung tissue, airway (over the lung tissue). A comparison of the performance of each application can be found in section 7.3.

6.4 Snakes

Snakes or active contour models [11], are an iterative algorithm to fit a parametric curve to image features by minimizing an energy function, usually based on the strength of the edges and curvature of the curve. This chapter describes the application of a slightly modified snakes based algorithm to the problem of airway segmentation.

The problem of airway segmentation does not immediately lend itself to a snakes based solution. The edges of the airways are relatively weak in comparison to the unrelated edges present at the edge of ribs and heart. The airway edges do not form a closed curve, or single line, but rather a set of lines fading into the background towards the lower branches.

However, the alignment of the training images (shown in Figure 7.2) suggests that edges of the airways will occur in similar places and in similar directions. This knowledge could be used to improve the separation between airway edges and background.

Another feature of the airways is that the opposing walls tend to lie in similar directions; in order to take advantage the snake was implemented as done by Golland et al. [33]. Two similar curves representing the edges of a single branch of the trachea were created, the midpoint between equivalent points on each curve was calculated and joined to the midpoints of its neighbours to form the "skeleton" of the trachea. Two identical structures were attached to the bottom end of these two curves to represent the left and right main bronchus. Figure 6.3 shows the model of the airway tree build using this method. The search was limited to only the trachea and main bronchi, with each section consisting of the three curves made up of 8 control points each. The model was initialized with a width of 0 pixels along the central axis of the probability map found in Section 6.2.1.

6.4.1 Energy Function

A single curve, $v(s)$, in the model can be represented by $v(s) = (x(s), y(s))$ where $xi(s), y(s)$ represent the coordinates of the control points forming the curve.The energy function, E_{snake}, for the entire model can be expressed as:

$$E_{snake} = \int_0^1 E_{image}(v(s)) + E_{LeftEdge}(v(s)) + E_{RightEdge}(v(s)) + E_{Skeleton}(v(s)) + E_{inflation}(v(s))ds$$

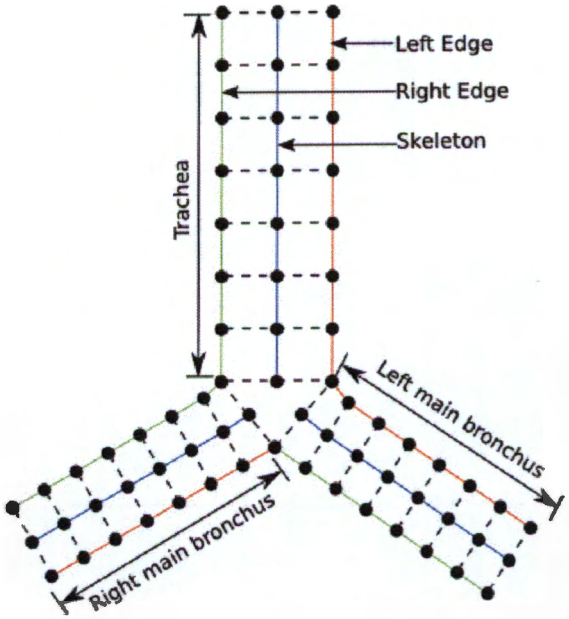


Figure 6.3: The model of the airway tree build from individual curves.

As in Kass et al.'s implementation [11], the energy function used for a single curve was,

$$E_{spline} = (\alpha(s)|v'(s)|^2 + \beta(s)|v''(s)|^2)/2$$

where the parameters $\alpha(s)$ and $\beta(s)$ effect the amount of curvature allowed in the snake, the same function and constants were used for both edges ($E_{RightEdge}$, $E_{LeftEdge}$) and the skeleton ($E_{Skeleton}$).

The energy due to image features, E_{image} was determined as the cross product of the differentiated image and a vector at right angles to the snake at the edge of the airway. The inflation pressure, $E_{inflation}(s)$, is calculated as a weighted sum of the distances between each equivalent point on the opposing sides of the model and is used to encourage the model to expand from its initialisation point and fill the whole airway. The most effective combination of weightings for each component was determined through trial and error, testing each variation on the training set of images.

Chapter 7

Results and Conclusions

7.1 Chapter Summary

The results of the airway segmentation are presented in this chapter. Conclusions and recommendations for future work are included in Section 7.5.

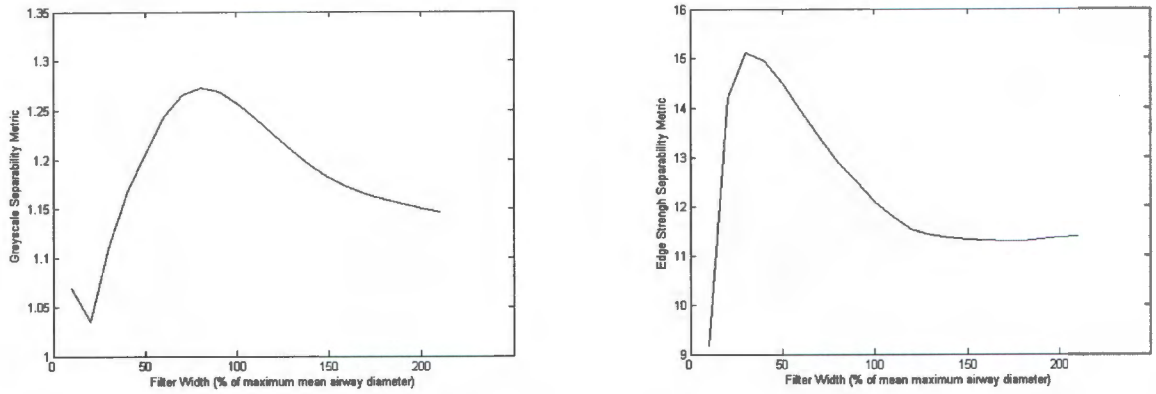
7.2 Pre-Processing

7.2.1 Filtering

Figure 7.1 shows the effects of changing the filter size on the strength of edges of the airway with respect to the average edge strength in the image, and on the airway interior average greyscale intensity with respect to average image intensity. The strength of edges was found to peak with a filter size of 35% of the average maximum airway diameter, while the internal greyscale intensity was lowest for a filter size of 80% of the maximum average airway diameter. Because of this, a filter size of 80% average maximum airway diameter is expected to provide the best results for segmentation techniques based on finding the darker pixels associated with the airways and so is used in the expectation maximization classification described in Section 6.3. A 35% average maximum airway diameter filter should provide the best images to be used in an edge based segmentation, so was used for the snake based algorithm discussed in Section 6.4.

7.2.2 Shape Normalisation

Figure 7.2 shows the effects of different warping methods on the resulting probability map constructed from the training images. The piece-wise linear warping including the first bifurcation point seems to provide the best alignment between the airways in the training set. This is not surprising given that this method utilises the largest amount of data gathered in the AAM search. This is confirmed by the results of the sharpness metric. Figure 7.3 shows the average energy of the image Laplacian for the probability maps built using each technique. It is expected that the piecewise linear warping method including the first bifurcation point will provide the best results from any segmentation algorithm, so the expectation maximization and snakes based segmentation techniques covered in Sections 6.3 and 6.4 both use images warped by this method.



(a) Greyscale separability

(b) Edge strength

Figure 7.1: The effects of changing the filter parameter on the separability metrics.

7.3 Expectation Maximization Classification

A subset of 20 images was selected from those used in Part One and the airways in these were manually segmented. Ten of these images were used as a training set for building the probability maps needed by the EM segmentation. The remaining ten were used as an unseen set with which to test the segmentation algorithms.

The expectation maximization algorithm was run on the test images after shape normalisation had been applied using the methods described in Section 6.2.1, and was found to converge to a stable solution after three iterations. To evaluate the effectiveness of applying the EM segmentation algorithm, the accuracy of classifications before and after running the algorithm were considered.

The airway shapes identified from the images using the EM classification algorithm can be seen in Figure 7.4, Figure 7.4a has had the 2 class EM classification applied whereas Figure 7.4b has been segmented using the 4 class method.

Because of the small size of the airways in comparison to the entire full-body x-ray, calculating the percentage of correctly labelled pixels can give misleading results; over 98% of pixels can be labelled correctly simply by labelling all pixels as non-airway. To avoid this, all calculations of segmentation accuracy are done ignoring the pixels labelled correctly as non-airway. For example, the percentage of correctly pixels 1abelled in Figures 7.5 and 7.8 is calculated as:

$$\frac{\textit{Number of pixels correctly labeled as airway}}{\textit{Number of incorrect pixels} + \textit{number of pixels correctly labeled as airways}}$$

Figure 7.5 shows the percentage of pixels correctly and incorrectly classified by the EM segmentation. The results shown in the figure are split into

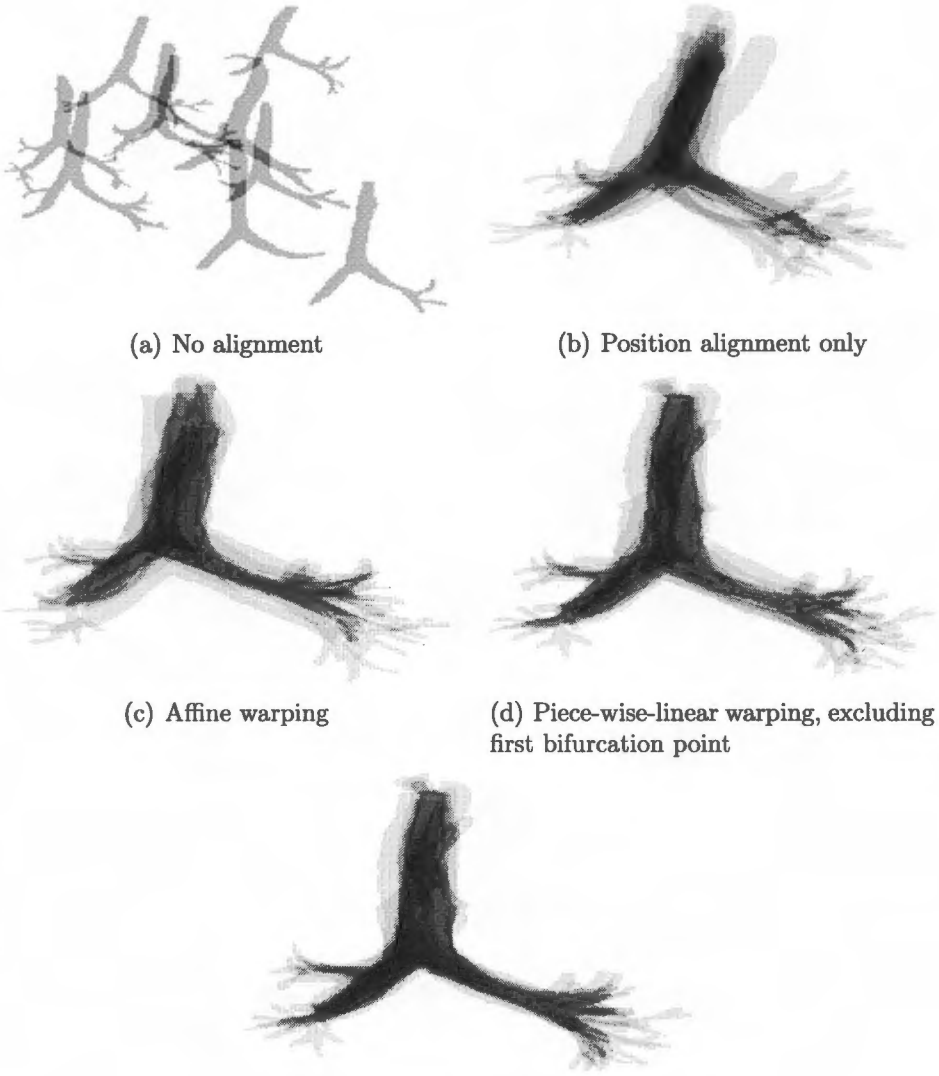
Correct - the percentage of airway pixels that were correctly labelled by the segmentation.

Incorrect - percentage of pixels incorrectly labelled by the segmentation.

False Negative - percentage of pixels belonging to the airways which were classified as background.

False Positive - percentage of pixels representing background which were incorrectly classified as airway.

In order to provide a reference point, the image was also classified based only on the probability map (without performing any iterations of the EM algorithm). These results are shown in



(a) No alignment

(b) Position alignment only

(c) Affine warping

(d) Piece-wise-linear warping, excluding first bifurcation point

(e) Piece-wise-linear warping, including first bifurcation point

Figure 7.2: Probability maps constructed using various methods of alignment.

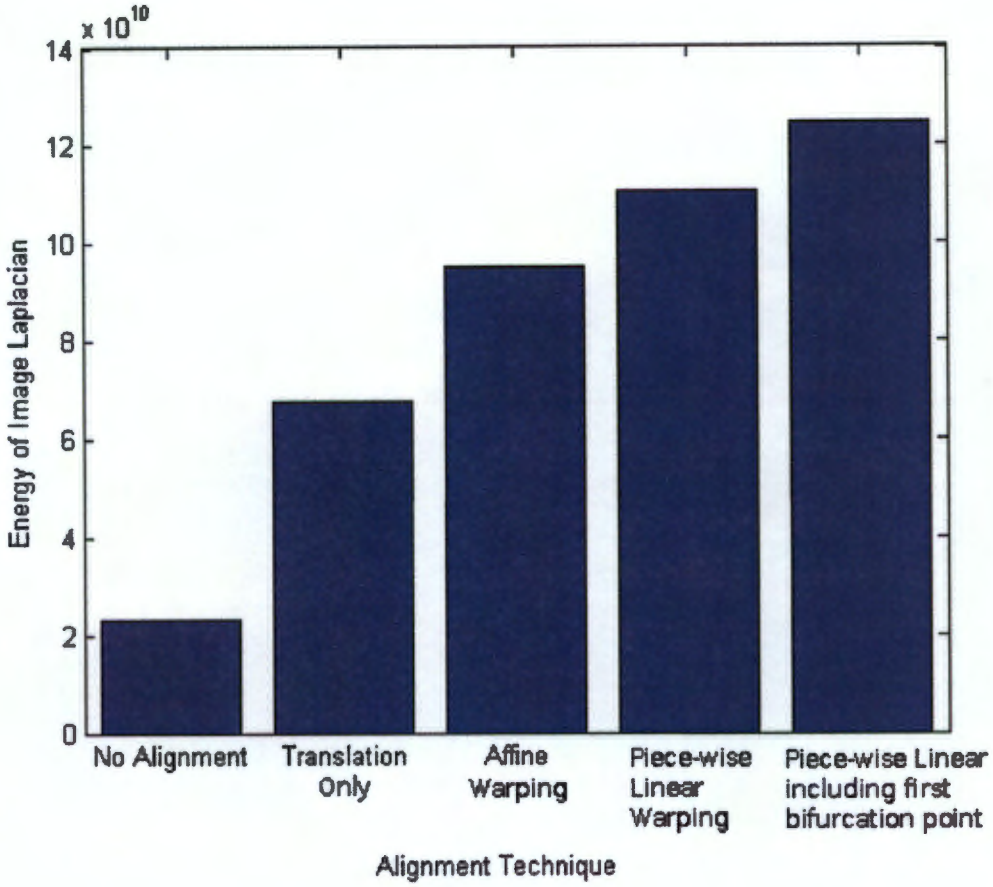
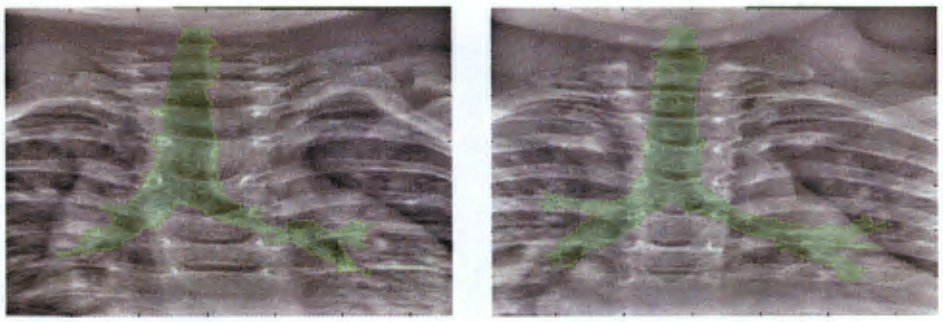


Figure 7.3: Sharpness metric for probability maps built using each method of alignment.



(a) 2 class EM classification. (b) 4 class EM classification.

Figure 7.4: Results of applying the EM classification to a shape normalised image of the airways.

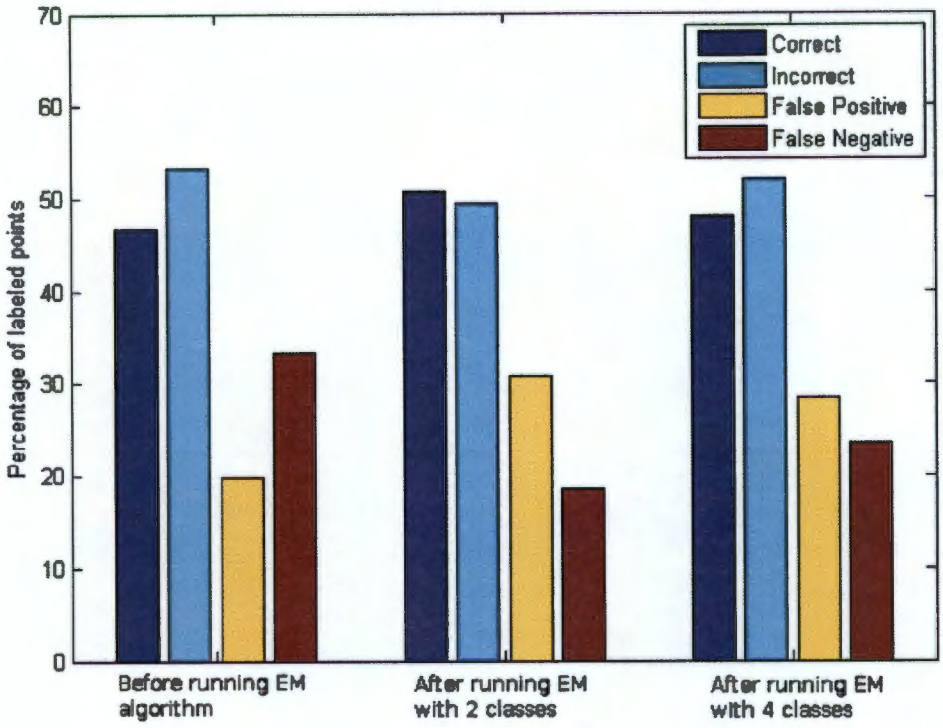


Figure 7.5: Results of EM classification.

Figure 7.5 as the first set of values. Classifying the image in this fashion is equivalent to taking the most likely airway shape and position and assigning it to each image without considering the actual content of the image. It is included here as a reference point to judge how much value is added by running the EM algorithm.

The segmentation using four classes proved to be slightly more effective than using two classes, resulting in 51% of pixels correctly identified compared to 48% when using two classes. However both two and four class classifications did not significantly improve upon the 46% of pixels correctly labelled by selecting the most likely airway position without running the EM algorithm.

In order to evaluate the effects of the shape normalisation on the EM classification, the classification algorithm was applied to the unseen images after they had been aligned using each of the techniques covered in Section 6.2.1. As expected, the piece-wise linear shape normalisation using the AAM search including first bifurcation point produced the most effective segmentation. Not aligning the images at all resulted in the worst performance, with none of the airway pixels being identified correctly. The effects of each technique of the overall accuracy correlates well with the predictions made by the image sharpness metric used in Section 6.2.1.

As expected, applying the EM classification to images without having applied the local normalisation filter (Section 6.2.2) produced a less accurate segmentation that when using the filter. EM classification applied to the unfiltered image set using the piece-wise linear shape normalisation produced 47% of pixels labelled correctly, compared to the 50% when using the filter. The effect of using unfiltered images in conjunction with the other alignment methods was not tested; however similar reductions in accuracy are expected.

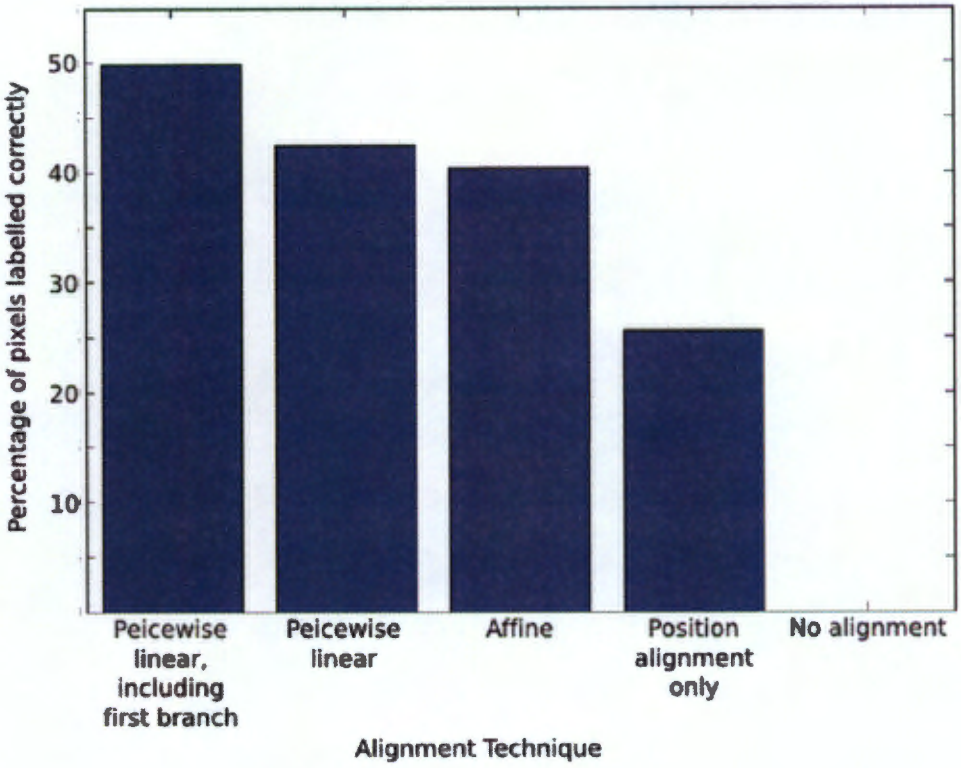


Figure 7.6: Effects of different shape normalization techniques on the effectivness of the EM classification

7.4 Snakes

The snakes algorithm applied to segmenting the airways, shown in Figure 7.7, did not perform as well as the expectation maximization classification. Despite the filtering done to try and enhance the airway edges, the algorithm was still prone to attaching to unrelated image features. Only 41% of pixels were correctly labelled using this method, less than the 46% obtained from selecting the most probably airway position (Section 7.3). Considering this, used on its own, the snakes algorithm is most probably insufficient to provide any useful segmentation of the airways.

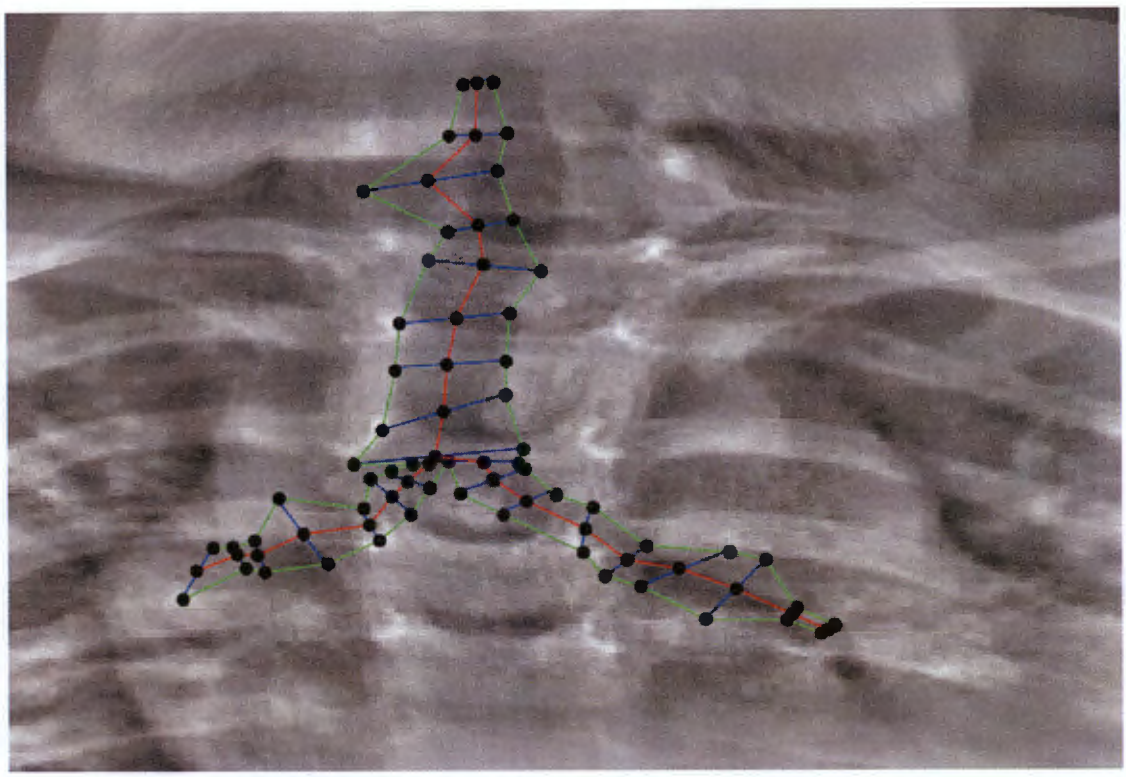


Figure 7.7: The results of a snake based search.

The snakes algorithm would be better suited to an application were a stronger edge in the image exists for the snake curve to attach itself to. This may possibly be used as an alternative to the edge tracing algorithm used to find the edges of the long bones in Section 3.2. Conversely it would also be possible to apply the bone edge tracing algorithm to the airway edges, however the same problems of weak edges and interference from stronger, unrelated edges would still apply.

Figure 7.8, shows the percentage of correctly labelled points for each of the segmentation algorithms tested. The snakes technique is clearly least likely to identify an airway pixel correctly, even when compared to segmentation based on the EM probability map without running the EM classification.

As predicted in Section 6.2.2, using the local normalisation filter resulted in a more accurate segmentation compared to not applying the filter: running the search on the unfiltered images reduced the percentage of correctly labelled pixels from 41% to 24%. However, as the snake parameters were selected based on a filtered training set, this is not a good indication of the effects of the filtering on the effectiveness of the algorithm. The effects of using different

alignment techniques was not considered as the initial position of the snake model was reliant on using the piece-wise linear alignment including first bifurcation point.

It maybe possible to improve the performance of the snake by performing additional pre-filtering steps to reduce the effects of background edges. Including more information about the expected shape of the airways could also produce a better segmentation; one method to accomplish this would be to use an active shape model instead of the snake curve.

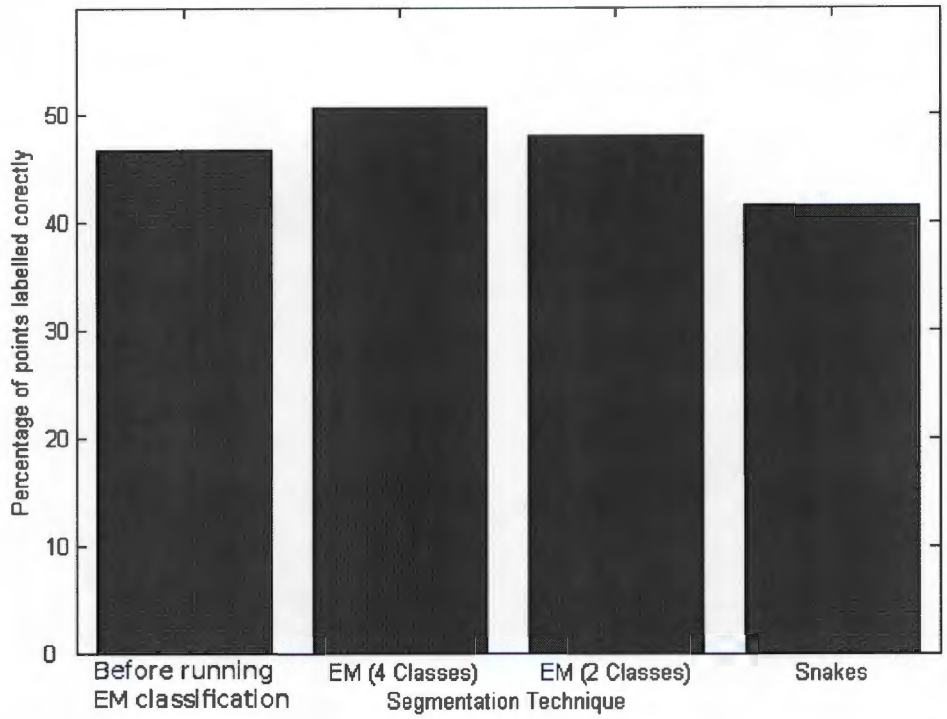


Figure 7.8: Comparison of implemented segmentation techniques.

7.5 Conclusions

Although shape normalisation did reduce the amount of variation in the training set, and the local normalisation filtering did improve the contrast of the airways, these improvements were not enough to provide reliable segmentations from either EM or snakes algorithms. As shown in Figure 7.8, none of the segmentation methods produce any significant improvement over classification based on the probability map alone and so are not expected to provide much diagnostic value.

In order to evaluate the performance of each method with regards to the different sections of the airway, the labels from the manually annotated images were used to identify which region the pixel belonged to. In the case of a false-positive pixel, it was considered to belong to the same class as the closest labelled pixel on the manual annotation. Figure 7.9 shows the results of this calculation.

It may be possible to improve the performance of both algorithms by using additional pre-filtering to suppress background features. Subtracting the bony features as attempted by

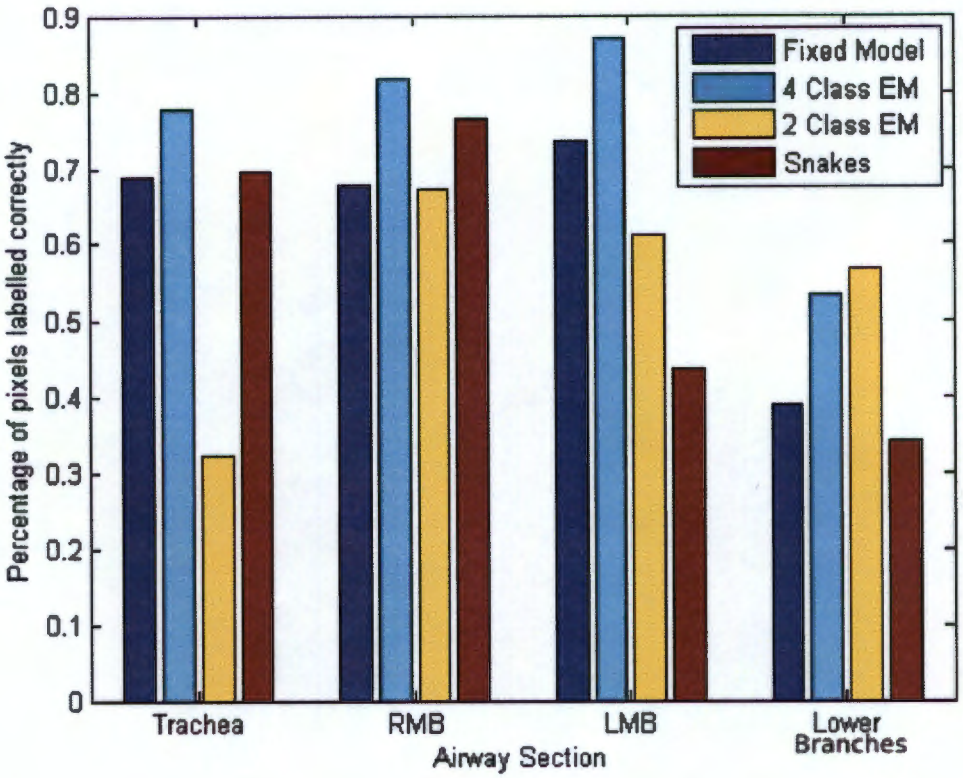


Figure 7.9: Performance of airway segmentation techniques with respect to different sections of the airway tree.

Schilham et. al. [34] and Suzuki et. al. [35] may be particularly useful provided that enough detail of the airways is preserved by the filtering operation.

Based on the poor performance of the tested techniques, a more feasible approach maybe to pursue a semi-automated segmentation aimed at improving the ease with which a human operator can segment the airways, rather than attempting a fully automated solution. Options in this regard include an edge tracing method initialized at human positioned seed points as used by Plourde et al. [8] to trace the rib boundaries. Human input could be incorporated into the implemented solutions if an operator were to trace the approximate centre line of the airway tree; this could be used as a more accurate initialisation point for the snakes or could be used to as a registration step for the EM segmentation probability map.

Although no testing was done in this regard, the pre-filtering methods discussed in Section 6.2.2 may also improve the visibility of the airways to a human operator. Further work into using these filters to assist human segmentation may be valuable especially if a semi-automated solution is to be considered.

References

[1] S. Beningfield, H. Potgieter, A. Nicol, S. van As, G. Bowie, E. Hering, and E. Lătti. Report on a new type of trauma full-body digital x-ray machine. *Emergency Radiology*, pages 23–29, 2003.

[2] RD. Pitcher, AB van As, V Sanders, TS Douglas, N Wieselthaler, A Vlok, S Paverd, T Kilborn, H Rode, H Potgieter, and SJ Beningfield. A pilot study evaluating the "statscan" digital x-ray machine in paediatric polytrauma. *Emergency Radiology*, 15:25–42, 2008.

[3] D. Biagio. Knoweldge based segmentation of full-body medical x-ray images. Master's thesis, University of Cape Town, 1996.

[4] G. Mori, X. Ren, A. Efros, and J. Malik. Recovering human body configurations: combining segmentation and recognition. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*, volume 2, pages 326–333. IEE, 2004.

[5] L. Li, Y. Zheng, M. Kallergi, and RA. Clark. Improved method for automatic identification of lung regions on chest radiographs. *Academic Radiology*, 8(7):629–667, 2001.

[6] B. van Ginneken, MB Stegmann, and M Loog. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Medical Image Analysis*, 10(1):19–40, Feburary 2006.

[7] MB. Stegmann. Active appearance models: Theory, extensions & cases. Master's thesis, Technical University of Denmark, 2000.

[8] F. Plourde, F. Cheriet, and J. Dansereau. Semi-automatic detection of scoliotic rib borders using chest radiographs. *Studies in Health Technology and Informatics*, 123:533–537, 2006.

[9] A. Koeslag. Computer aided diagnosis of millary tb in chest x-rays. Master's thesis, University of Cape Town, 2002.

[10] G. J. Edwards, T. F. Cootes, and C. J. Taylor. Face recognition using active appearance models. In H. Burkhardt and Springer-Verlag Bernd Neumann, editors, 5^{th} *European Conference on Computer Vision 1998*, volume LNCS-Series 1406–1607, pages 581–595, 1998.

[11] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[12] Z. Yue, A. Goshtasby, and L. Ackerman. Automatic detection of rib borders in chest radiographs. *IEEE Transactions on Medical Imaging*, 14(3):525–536, 1995.

[13] T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham. Active shape models– their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[14] M.R. Schilham, B. van Ginneken, and M. Loog. A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database. *Medical Image Analysis*, 10:247–258, 2006.

[15] M. Sonka and J.M. Fitzpatrick. *Handbook of Medical Imaging, Volume 2: Medical Image Processing and Analysis.* SPIE Press, 2000.

[16] M. B. Stegmann and D. D. Gomez. A brief introduction to statistical shape analysys. Technical report, Technical University of Denmark, DTU, March 2002.

[17] J. D. Foley, A. van Dam, S. K. Feiner, J. F. Hughes, and R. L. Phillips. *Introduction to Computer Graphics.* Addison-Wesley, 1993.

[18] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer vision. Technical report, University of Manchester, Feburary 2000.

[19] United nations site for the millenium development goals indicators. http://mdgs.un.org/, March 2006.

[20] Division of Tuberculosis Elimination US Dept. of Health. Core curriculum on tuberculosis. http://www.cdc.gov/nchstp/tb/pubs/corecurr/default.htm, 2000.

[21] Ejaz A. Khan and Jeffrey R. Starke. Diagnosis of tuberculosis in children: Increased need for better methods. *Emerging Infectious Diseases*, 1(4):115–138, 1995.

[22] D Sutton. *TextBook of Radiology and Medical Imaging.* Churchill Livingstone, 5 edition, 1993.

[23] C. Wei, C. Tiu, J. Chen, Y. Chou, C. Chang, and C. Yu. Computed tomography features of acute pulmonary tuberculosis. *American Journal of Emergency Medicine*, 22(3), May 2004.

[24] G. J Maree, B. J. Irving, and E. R. Hering. Paediatric dose measurement in a full-body digital radiography unit. *Paediatric Radiology*, 37:990–997, 2007.

[25] S. Andronikou, E. Joseph, S. Lucas, S. Brachmeyer, G. Du Toit, H. Zar, and G. Swingler. Ct scanning for the detection of tuberculosis mediastinal and hilar lymphandenopathy in children. *Pediatric Radiology*, 34:232–236, 2004.

[26] European Commission on Radiation protection 118. Referral guidelines for imaging. http://europa.eu.int/comm/environment/radprot/118/rp-118-en.pdf, 2001.

[27] E. Ron. Ionizing radiation and cancer risk: evidence from epidemiology. *Radiation Research*, 150(5 Suppl):30–41, November 1998.

[28] S. Sarkar and S. Chaudhuri. Evaluation and progression analysis of pulmonary tuberculosis from digital chest radiographs. *Computerized Medical Imaging and Graphics*, 22:145–155, 1998.

[29] J. Tschirren, E. Hoffman, G. McLennan, and M. Sonka. Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose ct scans. *IEEE Transactions on medical imaging*, 24(12):1529–1563, 2005.

[30] N. Moon, E. Bullit, K van Leemput, and G. Gerig. Model-based brain and tumor segmentation. In *Proceedings of the 16th International Conference on Pattern Recognition*, volume 1, 2002.

[31] A.M.R. Schilham, B. van Ginneken, and M. Loog. Multi-scale nodule detection in chest radiographs. In R.E. Ellis and T.M. Peters, editors, *Medical Image Computing and Computer-Assisted Intervention*, volume 2878 of *Lecture Notes in Computer Science*, pages 602–609. Springer, 2003.

[32] R. Gonzalez and R. Woods. *Digital Image Processing.* Addison-Wesley Longman Publishing Co., 2 edition, 1992.

[33] P. Golland, W. Grimson, and R. Kikinis. Statistical shape analysis using fixed topology skeletons: Corpus callosum study. *Inf Process Med Imaging.*, 16:382–387, 1999.

[34] M.R. Schilham, B. van Ginneken, and M. Loog. Filter learning: Application to suppression of bony structures from chest radiographs. *Medical Image Analysis*, 10(6):826–840, 2006.

[35] K. Suzuki, H. Abe, H. MacMahon, and K. Doi. Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (mtann). *IEEE Transactions on Medical Imaging*, 25(4):406–416, 2006.

Appendix

Table 7.1: The points used to build the simple (19 point) AAM model

Point Number	Position
1	Top Left Corner of Image
2	Top Right Corner of Image
3	Bottom Right Corner of Image
4	Bottom Left Corner of Image
5	Top of Head
6	Intersection of neckline and jaw (Right)
7	Intersection of neckline and jaw (Left)
8	Intersection of clavicle and scapula (Right)
9	Intersection of clavicle and scapula (Left)
10	Sterno-Clavicular Junction
11	Intersection of Humerus and scapula (Right)
12	Intersection of Humerus and scapula (Left)
13	Top of diaphragm on centerline of patient
14	Center of Elbow joint (Right)
15	Center of Elbow Joint (Left)
16	Center of Wrist (Right)
17	Center of Wrist (Left)
18	Right Hip Joint
19	Left Hip Joint
20	Right Knee Joint
21	Left Knee Joint
22	Right Ankle
23	Left Ankle

Table 7.2: The points used to build the more detailed (19 point) AAM model

Point Number	**Posiiton**
Point Number	Position
24	Top Right Corner of Skull
25	Top Left Corner of Skull

Continued on Next Page...

Table 7.2 – Continued

Point Number	Position
26	Ride Side of Skull in line with bottom of eye sockets
27	Left Side of Skull in line with bottom of eye sockets
28	Intersection of vertebrae of neck and jawbone (Right)
29	Intersection of vertebrae of neck and jawbone (Left)
30	Intersection of spine and top of rib cage (Right)
31	Intersection of spine and top of rib cage (Left)
32	Approximate point of transition between neck and shoulder (Right)
33	Approximate point of transition between neck and shoulder (Left)
34	Lateral most point on the head of the humerus (right)
35	Medial most point on the head of the humerus (right)
36	Top most edge of head of humerus (right)
37	Lateral most point on the head of the humerus (Left)
38	Medial most point on the head of the humerus (Left)
39	Top most edge of head of humerus (Left)
40	Lateral edge of neck of humerus (Right)
41	Medial edge of neck of humerus (Right)
42	Lateral edge of neck of humerus (Left)
43	Medial edge of neck of humerus (Left)
44	Intersection of Scapula and rib cage (right)
45	Intersection of Scapula and rib cage (left)
46	Lateral boundary of rib cage in line with the top of the diaphragm (Right)
47	Lateral boundary of rib cage in line with the top of the diaphragm (Left)
48	Lateral edge of humerus immediately distal to the epicondyle (Right)
49	Medial edge of humerus immediately distal to the epicondyle (Right)
50	Lateral edge of lateral epicondyle (Right)
51	Medial edge of medial epicondyle (Right)
52	Lateral edge of humerus immediately distal to the epicondyle (Left)
53	Medial edge of humerus immediately distal to the epicondyle (Left)
54	Lateral edge of lateral epicondyle (Left)
55	Medial edge of medial epicondyle (Left)
56	Lateral edge of radius at elbow joint (Left)
57	Medial edge of ulna at elbow joint (Left)
58	Medial edge of Wrist joint (Left)
59	Lateral edge of wrist joint (Left)
60	Tip of little finger (Left)
61	Tip of index finger (Left)
62	Lateral edge of radius at elbow joint (Right)
63	Medial edge of ulna at elbow joint (Right)
64	Medial edge of Wrist joint (Right)
65	Lateral edge of wrist joint (Right)
66	Tip of little finger (Right)
67	Tip of index finger (Right)
68	Lateral most point on the iliac crest (Right)
69	Lateral most point on the iliac crest (Left)

Continued on Next Page...

Table 7.2 – Continued

Point Number	Position
70	Lateral Intersection of femur and ilium (Right)
71	Lateral most point on the greater trochanter (Right)
72	Lesser trochanter (Right)
73	Lateral Intersection of femur and ilium (Left)
74	Lateral most point on the greater trochanter (Left)
75	Lesser trochanter (Left)
76	Lateral edge of femur at widest point above knee joint (Left)
77	Medial edge of femur at widest point above knee joint (Left)
78	Lateral edge of tibia at widest point below knee joint (Left)
79	Medial edge of tibia at widest point below knee joint (Left)
80	Lateral edge of femur at widest point above knee joint (Right)
81	Medial edge of femur at widest point above knee joint (Right)
82	Lateral edge of tibia at widest point below knee joint (Right)
83	Medial edge of tibia at widest point below knee joint (Right)
84	Lowest point of public symphysis
85	Lateral most point on ankle joint (Left)
86	Medial most point on ankle joint (Left)
87	Lateral most point on ankle joint (Right)
88	Medial most point on ankle joint (Right)
89	Tip of small toe (Left)
90	Tip of big toe (Left)
91	Tip of big toe (Right)
92	Tip of small toe (Left)