Automated Selection of Trabecular Bone Regions in Knee Radiographs

Published in:
Medical Physics

DOI:
10.1118/1.2905025

Link to publication in the UWA Research Repository

Rights statement
Post print of work supplied. Link to Publisher’s website supplied in Alternative Location.

General rights
Copyright owners retain the copyright for their material stored in the UWA Research Repository. The University grants no end-user rights beyond those which are provided by the Australian Copyright Act 1968. Users may make use of the material in the Repository providing due attribution is given and the use is in accordance with the Copyright Act 1968.

Take down policy
If you believe this document infringes copyright, raise a complaint by contacting repository-lib@uwa.edu.au. The document will be immediately withdrawn from public access while the complaint is being investigated.
Automated selection of trabecular bone regions in knee radiographs

P. Podsiadlo, M. Wolski and G. W. Stachowiak
Tribology Laboratory, School of Mechanical Engineering
The University of Western Australia, Crawley, Western Australia 6009

Abstract

Osteoarthritic (OA) changes in knee joints can be assessed by analysing the structure of trabecular bone (TB) in the tibia. This analysis is performed on TB regions selected manually by a human operator on x-ray images. Manual selection is time-consuming, tedious and expensive. Even if radiologist expert or highly trained person is available, to select regions, high inter- and intra-observer variabilities are still possible. A fully automated image segmentation method was, therefore, developed to select the bone regions for numerical analyses of changes in bone structures. The newly developed method consists of image pre-processing, delineation of cortical bone plates (active shape model) and location of regions of interest (ROI). The method was trained on an independent set of 40 x-ray images. Automatically selected regions were compared to the "gold standard" that contains ROIs selected manually by a radiologist expert on 132 x-ray images. All images were acquired from subjects locked in a standardized standing position using a radiography rig. The size of each ROI is 12.8×12.8mm. The automated method results showed a good agreement with the "gold standard" (similarity index (SI) = 0.83 (medial) and 0.81 (lateral) and the offset = [-1.78, 1.27]×[-0.65, 0.26] mm (medial)
and [-2.15, 1.59]×[-0.58, 0.52]mm (lateral)). Bland and Altman plots were constructed for fractal signatures and changes of fractal dimensions (FD) to region offsets calculated between the "gold standard" and automatically selected regions were calculated. The plots showed a random scatter and the 95% confidence intervals were (-0.006, 0.008) and (-0.001, 0.011). The changes of FDs to region offsets were less than 0.035. Previous studies showed that differences in FDs between non-OA and OA bone regions were greater than 0.05. ROIs were also selected by a second radiologist and then evaluated. Results indicated that the newly developed method could replace a human operator and it produces bone regions with an accuracy that is sufficient for fractal analyses of bone texture.

Key words: trabecular bone, computer-aided diagnosis, osteoarthritis, knee radiography
I. INTRODUCTION

Bone regions selected on digitized radiographs are used for the quantitative analysis of osteoarthritic (OA) changes occurring in knee joints. Previous studies showed that through the analysis of trabecular bone (TB) regions of the tibia OA changes can be detected in knees.\textsuperscript{1,3} Detection of OA is a key parameter in understanding the role of bone changes at the early stages of OA and during the disease progression. Inspite of its potential, the TB analysis is yet not used in routine clinical practice. The reason is that a human operator is required to select bone regions for the analysis first. Manual selection is time-consuming, tedious and expensive,\textsuperscript{3} and even if radiologist expert or highly trained person is available, to select regions, high inter- and intra-observer variabilities are still possible. The difficulties are caused by the fact that a large number of radiographs are analysed, the anatomy of a knee joint is complex and variable (e.g. OA knees can be severely deformed), the contrast between bone edges and background is often low, and special precautions are required to avoid selecting the periarticular osteopenia and the fibula head. For these reasons, an automated segmentation method that is fast, accurate and reproducible is highly desirable. In this study, such a segmentation method will be developed and used to select TB regions of interest (ROIs) in the tibia.

Our ultimate goal is to develop a fully automated system for the prediction, early detection and diagnosis of knee osteoarthritis. The automated selection of bone regions is a component of the automated system that is currently developed. Other components of the system include the texture characterization of bone regions, the measurement of joint space width, osteophytes and leg alignment. This would be a decision-support system that is
not meant to replace the radiologist, but when fully developed, it can improve the accuracy of predicting and finding early knee OA. As a result appropriate therapy can be started in a timely manner.

A large number of medical image segmentation methods have been proposed in the last 30 years, including thresholding, region growing, classifiers, clustering, Markov random field models, artificial neural networks, deformable models, atlas-guided approach and model-fitting. For example, active shape models (ASMs) were used to find a boundary of the femoral head in leg radiographs and to segment the tibia bone in ultrasound images, the posterior ribs were selected in chest radiographs using an iterated contextual pixel classification while hand radiographs were segmented by multiple connected active appearance models. However, none of these methods can be directly used to solve the problem of selecting bone regions in the tibia.

In this study, a new automated method for TB bone region selection has been developed. The method contains three components: (i) image pre-processing (fibula positioning, image resizing, contrast enhancement, delineation of the knee borders and the joint space line), (ii) delineation of the cortical bone plates (based on ASM) and (iii) location of ROIs (placing a 256×256 pixel region under the lateral and medial cortical bone plates). The performance of this method was tested on healthy and OA knee images using manually selected regions by a radiologist expert as the "gold standard".

II. METHODS

A. Data sets

1. Acquisition of radiographs
Plain anteroposterior radiographs of left and right knees in weight-bearing full extension were acquired from 86 subjects. The total number of radiographs is 172 (one per knee). Each individual subject was locked in a standarized standing position using a radiography rig. Focal calcaneus distance was 1 m and peak voltages were set between 55 and 65kVp. Exposure settings were adjusted for each subject to obtain an optimal visualization of the tibia bone. Radiographic films were digitized using a Linotype Saphir Ultra 2 scanner with a scan mode of 16-bit RGB and an optical resolution of 500dpi. Radiographs were cropped to 2048×2048 pixels during the scanning with the purpose of reducing the storage space of X-ray database. To reduce computational times and to obtain smooth images, digitized radiographs were converted to 8-bit gray scale level images with a pixel resolution of 0.05 by 0.05mm. Previous studies showed that this spatial resolution is sufficient for the texture analysis of bone regions.

2. Normal and OA knees

All subjects were diagnosed according to the following criteria for radiographic knee OA:

(i) a joint space narrowing (JSN) of grade 2 or higher,
(ii) the sum of the 2 marginal osteophyte grades from the same compartment equal or greater than 2, or
(iii) a grade 1 JSN in combination with a grade 1 osteophyte in the same compartment.
The criteria approximate a grade 2 knee OA based on the Kellgren/Lawrence scale.¹¹ For the grading of knees the Osteoarthritis Research Society International (OARSI) atlas was used. 15 (out of 172) knees were identified as having the radiographic OA.

3. "Gold standard"

The performance of the automated selection method developed was evaluated by comparing the results obtained against a "gold standard". The standard contains 344 ROIs (two 256×256 pixel ROIs per knee image) selected by a radiologist experienced in the selection of bone regions. As an example, manually selected regions are shown in Fig. 1. To speed the selection process, a custom mouse pointer tool was used to assist the radiologist in performing image operations. First, the radiologist used a computer mouse to click on the most outer points of the lateral and medial tibia borders (osteophytes were excluded). The pointer tool was then used to automatically execute the following steps:

(i) drawing a vertical line through the middle between the outer points,
(ii) locating a region point in the horizontal distance set to the 1/3 of the compartment width measured from the lateral or medial outer point to the central line. The one-third width was used to avoid the periarticular osteopenia adjacent to marginal osteophyte formation,
(iii) drawing and centering a ROI at the region point and then moving it horizontally toward the knee center.
Next, the radiologist moved vertically the region until its top boundary was located immediately under the medial or lateral subchondral sclerosis area. Finally, the radiologist moved horizontally the region toward the central line to avoid an overlapping with the fibula head. Throughout this process the radiologist had to adjust manually the image brightness, contrast and magnification because of low contrast between the bony structure and soft tissue. The radiologist had to use edge detection filters since the sclerosis area and the fibula were difficult to identify. The selection of ROIs was repeated independently by a second radiologist. The rationale behind the use of two radiologists was an assessment of inter-observer variability in manually selected regions.

The first radiologist also used a computer mouse to click on 10 points along a part of the fibula head border that overlaps the tibia bone. These points are joined together by lines and a curve that is an approximation of the head border is obtained. This curve will be used in the calculation of a similarity index measure, i.e. the amount of false positive pixels (to be defined later).

For the construction of the "gold standard" rules were developed on the basis of previous works in the area of bone texture analyses.\textsuperscript{1,2,12} For example, each bone region was sized to 256×256 pixels which cover an area of 12.8×12.8 mm. Similar area sizes were used in previous studies.\textsuperscript{2,3,13-15} To avoid the periarticular osteopenia, the 1/3 compartment width measured from the lateral (medial) point to the center line was used. This width is similar to the outer 1/4 of the tibial compartment width used by Messent et al.\textsuperscript{2}

4. Training and testing data sets
172 knee images were split into training and testing data sets. The training set contains 40 knee images (i.e. 4 OA knee images and 36 images obtained from the normal cases). This set of images was used to construct an ASM of the cortical bone plate and to find the optimal thresholds for the automated method. 132 remaining images were used to build the testing set.

B. Description of method

1. Image pre-processing

a. Fibula position. X-ray images were read from the database along with the information whether they represent the left or right knee. Based on this information, images containing the right knee were mirrored horizontally. This produced images that have the fibula head located at the right side.

b. Image enhancement. Each image was sized down to 512×512 pixels from the original size of 2048×2048 pixels using bicubic interpolation and its 256 gray scale levels were rescaled to range from 0 to 1. The image contrast was adjusted using a contrast-limited adaptive histogram equalization (CLAHE) method combined with a median filtering and a Sobel edge sharpening. This image processing produced sharp, well defined edges of the tibia and femur bones; as shown in Fig. 2(b). A median filter was used to reduce the amount of high-frequency noise in the image.

c. Delineation of knee borders. Knee borders were determined on the lateral and medial sides in the following three steps:

(i) A morphological dilation operation with the length of an 8×1 pixel vertical structuring element was applied five times to the image. The element size and the number of applications have been empirically
determined based on all cases in the training set. These values were sufficient to close the joint space between femur and tibia bones. (ii) The image dilated was binarized. The white color represents bone regions and non-anatomical features such as a ball bearing used for scale calibration and labels used as markers on radiographic films. The non-anatomical features were eliminated by area sizes since they were smaller than bone regions. (iii) The lateral and medial bone borders were found as side edges of the corresponding bone regions [Fig. 2(c)].

\textit{d. Determination of the “joint space line”}. An integral projection method\textsuperscript{17} was first applied line by line in the horizontal direction to a rectangular region selected on the original knee image [Fig. 2(a)] and then, to a rectangular region selected on the enhanced image [Fig. 2(b)] that was transformed by a Mallat horizontal wavelet. Vertical borders of each region were located on the mean vertical lines of knee borders [Fig. 2(c)]. The region height was equal to the image size. The projection method produced a horizontal line for each region. The two lines were located at the positions corresponding to the largest gradient (i.e., the biggest change in gray level values). The “joint space line” was defined as the mean between these lines [Fig. 2(d)]. The largest gradient was used because the largest variation in gray level values was observed between the proximal tibia (or distal femur) edge and the joint space and between the cortical bone plate and its surrounding. The mean line was used to find the last landmark point for the initialization of ASM (to be shown latter).

2. Delineation of cortical bone plates
An ASM method was used to find a shape that best fits the superior border of the medial or lateral cortical plates. The method was chosen because it works well with x-ray images in the presence of a low contrast between overlapping objects, between the object and its surrounding area and has the ability to capture the complexity and inter-individual variability of the knee shapes. The shape fitting was done in the following manner:

\textit{a. Landmark points.} First, a training set of shapes of the superior border of the cortical plate was found manually by a radiologist in $N_t$ knee images [Fig. 3(a)]. A horizontal Mallat's 2D wavelet transform\textsuperscript{21} was used to enhance image features in horizontal directions. Each shape can be described by a shape vector $\mathbf{S}_i = [x_{ij}, y_{ij}, \ldots, x_{in}, y_{in}]^T, i = 1, 2, \ldots, N_t$, where $x_{ij}, y_{ij}$ are the $x, y$ coordinates of the $j$-th landmark point in the $i$-th shape and $n$ is the number of landmark points.

\textit{b. Shape alignment and statistics of aligned shapes.} Each shape vector was then aligned with respect to a set of axes using a modified Procrustes alignment algorithm.\textsuperscript{20} These aligned shapes were used to calculate the mean shape model $\overline{\mathbf{S}}$ and the covariance matrix $\mathbf{C}$, i.e.

$$\overline{\mathbf{S}} = \frac{1}{N_t} \sum_{i=1}^{N_t} \mathbf{S}_{ai} \quad \text{and} \quad \mathbf{C} = \frac{1}{N_t} \sum_{i=1}^{N_t} (\mathbf{S}_{ai} - \overline{\mathbf{S}})(\mathbf{S}_{ai} - \overline{\mathbf{S}})^T$$

where $\mathbf{S}_{ai}$ is the $i$-th aligned shape.

\textit{c. Shape model.} Every shape in the training set was approximated by the mean shape and a linear combination of eigenvectors $\mathbf{S} = \overline{\mathbf{S}} + \Phi \mathbf{b}$, where $\Phi = [\Phi_1, \Phi_2, \ldots, \Phi_m]$ are the eigenvectors of the covariance matrix $\mathbf{C}$; corresponding to the $m$ largest eigenvalues and $\mathbf{b}$ is a vector of $m$ shape parameters. The desired number of modes $m$ was defined as the smallest
number that satisfies the condition \[ \sum_{i=1}^{m} \lambda_i \geq f_v \sum_{i=1}^{2n} \lambda_i, \] where \( \lambda_i \) is the \( i \)-th largest eigenvalue of \( C \) and \( f_v \) is a portion of the variance in the training set. The parameter \( f_v = 0.9 \) was used in this study. The approximation of shapes from the training set was performed by varying the shape parameters \( b \) within 
\[-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i}, \quad i = 1, 2, \ldots, m.\]

d. Gray-level appearance model. The appearance model was used in model fitting required for detecting the cortical plate border. It describes the typical image feature profiles around each landmark point. To construct the model, first an image profile was centered at each landmark point, perpendicular to the contour. The normalized first derivatives of the profile were then calculated for each landmark point. This process was repeated for every image taken from the training set. Finally, using the first derivatives obtained for all images the mean profile \( g_j \) and the covariance matrix \( C_w \) were computed for each landmark point \( (j = 1, 2, \ldots, n) \).

e. Model fitting. Having generated the appearance model a search of the cortical plate bone border was performed. This involved finding the pose parameters \( p = (t_x, t_y, s, \theta) \) and the shape parameter \( b \) such that the ASM \( S = A(p) \cdot [\bar{S} + \Phi b] \) best fits target points that are most likely to be the cortical bone border. \( t_x, t_y \) represent translations in \( x - y \) directions and \( s, \theta \) are the scaling and rotation transformations.

The model fitting process was performed using following hypotheses. First, it was hypothesised that the translation, rotation and shape parameters are equal to zero and the scale is 1, and all landmark points of the mean shape model were set in such a way that the last landmark is located on the cross point of the knee border and the "joint space line" [Fig. 3(b)]. To fit the model,
the hypothesised parameters were adjusted by finding the minimum Mahalanobis distance calculated between a new image feature profile $g_{\text{new}j}$ and the mean profile $g_j$

$$d^2(g_{\text{new}j}) = (g_{\text{new}j} - g_j)^T C^{-1}_{g_j} (g_{\text{new}j} - g_j).$$

For this purpose, five new profiles were generated, i.e., the profile $g_{\text{new}j}$ centred at the $j$-th current landmark point ($j = 1, 2, ..., n$), and four profiles obtained by moving 5 and 10 pixel up and down $g_{\text{new}j}$ in the normal direction to the curve passing the landmark point. The minimization process resulted in the first shape model and the first Mahalanobis distance calculated. In the remaining steps, the following hypotheses were used for the parameters: scales with values of 0.9 and 1, the rotation equals 0 and the translation in vertical direction from -10 to 60 pixels in steps of 5 pixels determined with respect to the "joint space line". For each hypothesis, both the shape model and the Mahalanobis distance were calculated. The distances were compared and the shape model with the smallest distance was selected. This model was used to produce landmark points that represent the cortical bone plate.

### 3. Location of ROIs

Separate ROIs were selected on the lateral and medial regions of the tibia. Each region was initially placed in such a way that the middle point of the upper boundary was immediately under the lowest landmark point of the cortical plate (Fig. 4).

*a. Horizontal adjustment.* The horizontal position of ROI was adjusted to account for a periarticular osteopenia adjacent to marginal osteophyte
formation and osteophyte growth on bone borders. This was achieved as follows. First, mean lines were constructed by averaging data points in the x-direction for the lateral or medial border lines. A central vertical line running in the middle between these mean lines was then constructed. Next, a reference point was defined at the horizontal distance set to the threshold 1/3 of the compartment width measured from the lateral or medial mean value of border line to the central line. The region was centered at the reference point and finally, it was moved horizontally towards the knee center by the distance equal to a half of the region size [Fig. 5].

b. Vertical adjustment. The vertical position of ROI was adjusted to account for a subchondral bone sclerosis; as shown in Fig. 6. A rectangular region (the dotted line region) was constructed using ROI (the solid-line region). First, the ROI was enlarged by 1/8 and 1/4 in the horizontal and vertical directions, respectively and was then moved vertically by the distance equal to 1/8 of the ROI height. The rectangular region resulted contains subchondral bone sclerosis. This is because the ROI was selected immediately under the inferior border of the cortical plates where the bone sclerosis occurs.2 Horizontal features within the image region were then enhanced using a Mallat horizontal wavelet transform.21 Next, the enhanced image was thresholded. This produced a binary image with several white regions. Finally, a white region with the largest area was identified. This region represents the subchondral sclerosis area and the ROI was moved vertically under it.

c. Final adjustment. The horizontal position of the lateral ROI may require a further adjustment to avoid an overlapping with the fibula head. This was achieved by fitting an ellipse to the fibula head, and by moving the ROI in the horizontal direction towards the knee center until the fitted ellipse and the ROI
did not overlap. An ellipse was fitted into two points representing the superior aspect and the medial aspect of the fibula head. The center of the ellipse (marked by the star shown in Fig. 7(j)) was located at a cross-point of the horizontal and vertical lines running through the medial and superior aspect of the fibula head, respectively. The superior aspect was defined as a cross-point between the lateral knee border and the horizontal line running through the middle of ROI. This point is represented by the solid dot shown in Fig. 7(a). The other point was obtained by executing the following steps:

(i) A rectangular region of the original x-ray image was selected in such a way that the upper boundary runs through the middle of the ROI and the medial boundary runs through the middle of the tibial bone [Fig. 7(b)]. The region selected was resized to a 256×256 pixel image.

(ii) First, the contrast of the square image was adjusted using the CLAHE method [Fig. 7(c)]. A Mallat’s 2D wavelet transform was then applied to the image and modulus images were obtained. The original image was reconstructed at scales $2^2$ and $2^4$ without horizontal components. White pixels indicating the edge curves that have the highest amplitude values [Fig. 7(d)]. Using thresholding a binary image was obtained [Fig. 7(e)]. The white region indicates an edge of the tibia.

(iii) The step (ii) was repeated, but in addition to the CLAHE method a 9×9 median filter was used and the image reconstruction was performed at scales $2^1$, $2^2$ and $2^3$. The resulting image [Fig. 7(f)] was thresholded and binarized [Fig. 7(g)]. The white regions represent areas of the most prominent edge curves and textural features of the tibia and fibula in vertical directions.
(iv) Using binary images [Fig. 7(e) and 7(h)] an entry point between the tibia and the fibula head was found. The entry point was defined as the superior aspect of the largest white region in Fig. 7(h); located with respect to the lateral site of the white region shown in Fig. 7(e). This point is represented by the cross shown in Fig. 7(i).

(v) A part of the largest region of the binary image shown in Fig. 7(i) that lies below the cut-off horizontal line was removed. This produced a white region that covers a medial part of the fibula head [Fig. 7(j)]. The medial aspect of the fibula head [the cross in Fig. 7(j)] was determined by locating the most medial point of the white region.

Figure 7(k) shows an ellipse fitted into the fibula head and a bone ROI. The horizontal position of the ROI was adjusted so that there is no overlap with the ellipse and subsequently, with the fibula head.

C. Evaluation measures

a. Similarity index and offset. The performance of the automated bone segmentation method was evaluated versus the "gold standard" using a similarity index (SI) measure. The SI is an overlap measure calculated between the automatically selected ROI and the corresponding ROI taken from the "gold standard" [Fig. 8]. The measure was defined as:

\[ SI = \frac{2 \times TP}{2 \times TP + FPa + FPb^2 + FN} \]

where TP and FN represent the amount of true positive (correctly selected pixels) and false negative pixels (pixels selected as background, but in reality
belonging to the gold standard), respectively. FPa and FPb are the amount of false positive pixels, i.e. pixels selected as the gold standard, but in reality belonging to (a) background or (b) the fibula head. The measure takes values between 0 and 1. A high value represents a better agreement with the "gold standard". The measure attains the maximum value 1 if the selected ROI completely covers the gold standard ROI. It is equal to 0 if those ROIs are disjoint.

When calculating the SI measure, a square of FPb was used to ensure that an overlapping with the fibula head is heavily penalized. The penalty is necessary because values of fractal dimensions (FDs) are affected by the presence of the fibula head in ROI (to be shown later).

The performance of the automated method developed was also evaluated using an offset distance in the x and y direction. The distance is given in millimeters and one pixel corresponds to 0.05mm. Offsets were used because they are easy to calculate and provide an accurate measure of distances between square ROIs in the x and y directions.

b. Fractal signatures. ROIs were selected for numerical analyses of changes in bone structures. For these analyses, fractal texture descriptors have been found particularly useful.\textsuperscript{1-3} One important descriptor is a fractal signature (FS),\textsuperscript{1} i.e. FDs at individual TB widths. Previous study showed that this descriptor provides the most precise characterization of TB texture.\textsuperscript{1,3,10} It is, therefore, important to assess an agreement between FSs calculated for corresponding regions obtained from the "gold standard", the automated method and the second radiologist. For this assessment, a Bland and Altman plot was used.\textsuperscript{24} The plot represents the difference versus average of FDs calculated for the "gold standard" and the automated method (or the second radiologist) at individual
TB width. A 95% confidence interval (i.e. the limit of inter-reader agreement) around the mean difference between corresponding FDs was used. FDs were calculated at each TB width in the range of 0.3 to 0.65mm in steps of 0.05mm.

The effect of region offsets on FSs was evaluated. This was done by comparing FSs calculated for a "gold standard" region with those calculated for the region that was offset in x and y directions. The effect of overlapping between ROI and the fibula head was also evaluated.

III. RESULTS

A. Finding optimal thresholds

Binary images were used to determine the vertical and horizontal positions of the bone ROIs. Threshold values used to obtain the images and value of another threshold, i.e. the 1/3 of the compartment width used to avoid the periarticular osteopenias, were initially chosen by a human operator. Since the accuracy in positioning of the ROIs depends on these thresholds a wrong selection of their values can lead to errors. The SI measure, therefore, was used to find optimal thresholds. An exhaustive search was performed for wide ranges of threshold values. All possible combinations of thresholds were not used to find optimal values since this would require prohibitively long computational times. For each threshold value, SI was calculated between automatically selected ROIs and the gold standard ROIs taken from the training set, and the following three plots were constructed:

(i) Medial compartment plot [Fig. 9(a)]: For ROIs selected in the medial compartment, SI measures were plotted against two thresholds. The
first threshold (Threshold 1) produced binary images used for the vertical position adjustment, while the second threshold (Threshold 2) was used to avoid the overlapping with a periarticular osteopenia. Threshold 1, which is the image grey level value, was changed from 0.4 to 1, in steps of 0.05. Threshold 2 represents a fraction of the medial compartment width and its values were changed from 0.21 to 0.41, in steps of 0.02.

(ii) Lateral compartment plot [Fig. 9(b)]: This plot is the same as the previous one, but constructed for the lateral compartment. Thresholds 3 and 4 were used in the place of Thresholds 1 and 2.

(iii) Fibula head plot [Fig. 9(c)]: This plot is constructed to find optimal thresholds for an ellipse fitting into the fibula head. Thresholds 5 and 6 were grey level values and they were changed from 0.25 to 1, in steps of 0.05.

Using the plots optimal thresholds were selected by finding the maximum value of the SI measure. The thresholds selected were 0.75, 0.29 [Fig. 9(a)], 0.90, 0.29 [Fig. 9(b)] and 0.45, 0.65 [Fig. 9(c)], respectively.

B. Evaluation results

The performance of the automated method developed was evaluated on the testing set of 132 images. SI and offsets were calculated between automatically selected and the "gold standard" ROIs and between selected by the second radiologist and the "gold standard" ROIs. Mean, 95% confidence range and interval values of all measures calculated are listed in Table I.

Bland and Altman plots were constructed for FSs, and the plots together
with 95% confidence intervals are shown in Fig. 10.

FSs were calculated in the vertical and horizontal direction for each "gold standard" region and ROIs that were obtained by offsetting the "gold standard" region, in steps of 0.05mm. Values of the offsets used lie within a rectangle of [-3,3]×[-1,1]mm. The rectangle size was chosen to ensure that offset values listed in Table I were included. Mean values of the changes of FSs between the "gold standard" regions and the offsets were calculated. It was found that FDs calculated at scale of 0.65mm exhibit the largest changes with offset. The mean values calculated at this scale were thresholded and represented as a binary image [Fig. 11]. The threshold value was set to 0.035. This value represents the maximum change in FDs for offsets lying within the 95% confidence intervals (Table I).

FSs were also calculated for a ROI that was overlapped by fibula head. A percentage area of the overlapping is ranging from 0% to 15% in steps of 1%. As an example, FDs calculated at the scale of 0.65mm for increasing area of overlap are shown in Fig. 12.

C. Worst ROI selection results

The automated method can produce regions that are visibly different from the "gold standard" regions. Examples of poorly selected regions (white) and the "gold standard" regions (black) are shown in Fig. 13. SI and x-y offset measures calculated for these regions were low (high), i.e. 0.56 (-5.73mm, 0 mm), 0.02 (-2.04 mm, -0.46mm) and 0.16 (-10.75mm, -0.18mm). The second SI takes the small value of 0.02 because of ROI overlaps the fibula head. For comparison, measures obtained for the second radiologist were 0.95 (-0.55mm, 0.16mm), 0.90 (-0.88mm, 0.46mm) and 0.88 (-0.98mm, 0.69mm). Major errors in
selecting these regions can be associated with incorrectly identified knee borders [Fig. 13(a)] and inaccurately fitted ellipses that underestimated [Fig. 13(b)] or overestimated [Fig. 13(c)] the fibula head.

D. Human and computational times

The manual selection of a single ROI was performed at a rate slightly longer than 1 minute. Average time required by the radiologist for selecting 30 landmark points of the cortical plate border (done only for the training cases) took about 1.5 minute. The computational time required for one-time ASM training per region was about 1 minute. The automated selection of a single region took approximately 0.5 minute. All computer programs were written in Matlab and they were executed on an Unix computer with 1.2GHz clock. Times required for training the automated method and selecting a new ROI were listed in Table II.

IV. DISCUSSION

In this study, a fully automated method was developed to select ROIs of trabecular bone in knee radiographs. The method was trained on an independent set of 40 images and evaluated on a set of 132 test images. Two ROIs, i.e., the lateral and medial region, were selected on each image. The SI and offsets were calculated between the automatically selected ROIs and the "gold standard" regions. The SI measures were greater than 0.8. The SI measures obtained for the second radiologist were greater than 0.9. This indicates that the manual selection is more accurate than the automated method. However, results obtained with the automated method might improve when thresholds used are refined by increasing the number of training
examples.

The accuracy of the newly developed method for numerical analysing TB was evaluated using FDs calculated for the "gold standard" regions and regions selected by the automated method. Bland and Altman plots constructed for FDs calculated exhibit a random scatter [Fig. 10(a,b)], i.e. they do not vary in a systematic way over the range of average FDs. 95% confidence range and intervals were calculated [Fig. 10(a,b)]. Their values are small enough to be confident that the automated method can be used instead of the manual selection of regions performed by a radiologist expert. The results are comparable to those obtained for the second radiologist [Fig. 10(c,d)].

Changes of FDs to region offsets were calculated. Figure 11 shows that differences in FDs were no greater than 0.035 for most of the offset values obtained between the "gold standard" and automatically selected regions. Previous study\(^3\) showed that differences in FDs between non-OA and OA bone regions were greater than 0.05. This indicates that the automated method developed has potential for use in detecting TB changes in OA knee joints.

Bone ROIs can be overlapped by the fibula head. Figure 12 shows that FD is significantly affected by the overlapping area (>1%). In this study, the penalty term FPb in SI was used to minimize the overlapping area [Fig. 8]. Consequently, FPb minimizes the effect of the overlapping on FD.

The "gold standard" was constructed from bone regions that were selected by the radiologist expert who used the custom mouse pointer tool. The rationale behind the use of the pointer tool was that shifting the ROIs, placed automatically in reference to the radiologist-identified landmarks, would be easier than direct manual selecting ROIs (i.e. the selection of ROIs without the use of the pointer tool). It might be argued that the "gold standard" is bias, since
some of the rules used to place the ROI in reference to the landmarks are similar to those used in the automated method. However, where the "gold standard" region is finally placed depends entirely on the radiologist decision.

There are problems associated with the selection of bone regions. The "gold standard" was constructed from bone regions that were manually selected by a single human reader. Although the radiologist expert selected the bone regions, the interpretation and measurement errors may occur in the construction of this standard. The first type error occurs mainly along the knee border, subchondral sclerosis and fibula head that might be difficult to interpret by the human reader. Misinterpretations are more likely to occur when there is a low intensity contrast between the knee and the surrounding background and the bones of a knee joint are deformed. The second type of error is associated directly with the fact that clicking on the points along the knee borders and adjusting ROIs take long time and these are the eye-straining tasks. A possible way to minimize these errors is through the use of multiple readers.\textsuperscript{25} However, for a large database such as the one used in this study, this is impractical as each reader would have to analyze the whole database.

The "gold standard" was constructed from bone regions that were selected by the radiologist expert who used the custom mouse pointer tool. The rationale behind the use of the pointer tool was that shifting the ROIs placed automatically in reference to the radiologist-identified landmarks would be easier than direct manual selecting ROIs (i.e. the selection of ROIs without the use of the pointer tool). It might be argued that the "gold standard" is bias, since some of the rules used to place the ROI in reference to the landmarks are similar to those used in the automated method. However, where the "gold standard" region is finally placed depends entirely on the radiologist decision.
Another problem may arise if the lower part of the cortical plates is thinner and smaller as compared to the top part. In this case, the ASM may have difficulties with capturing cortical bone plates accurately. This problem can be overcome by using a statistical model constructed at different spatial scales. However, the process of constructing such a model requires a large training set and takes long time.25

Results obtained for ROIs selected on the lateral site were worse than those for the medial site (Table I). The worse performance can be associated with overlapping between the tibial bone and the fibula head. The head was sometimes not clearly visible and this influenced the ellipse fitting as shown in Fig. 13.

Good technical quality images of normal knee joints and few images of OA knees with deformed bones were used in this study. The images were obtained from knee radiographs of subjects who were locked in a same standardized position. Testing on poor quality images and poorly positioned subjects can give worse results. However, as the automated method is trainable, it can be tailored to different patient positions and measurement conditions such as image noise and exposure. This ability would be useful since it was found that FD does not change significantly with projection angle, image noise (<5%) and exposure.26

Some ROIs were incorrectly selected using the automated method. Since it is necessary to retain all data from a clinical study, a user interface was developed to adjust these ROIs manually. Results showed in this study were produced by the automated method without any manual adjustment.

Generally the automated method produces ROIs that are not as close to the "gold standard" as those provided by the human expert. This is consistent
with results reported for other automated methods, e.g. automated segmentations of lung fields in chest radiographs.\textsuperscript{27} It is not anticipated that the automated method will produce accurate region selections in all cases. In clinical trials, radiologists select bone regions from a large number of x-ray images. This manual selection is more expensive and time consuming than automated selections, even if the reader uses semi-manual selection tools and some results obtained from the automated method must be manually corrected. Furthermore, it cannot be ensured that follow-up studies will be performed by the same reader or the reader with the same level of experience.

Methods developed for the automated bone region selection can be modified and used in other applications. For example, in the measurement of a joint space width a manual image processing intervention is required to eliminate nonanatomical structures and to find a center of the knee joint and boundaries of the medial and lateral compartments.\textsuperscript{28} These image processing tasks can be automated using the knee border delineation methods. In the segmentation of tibia from knee x-rays\textsuperscript{29} the starting point used for ASM can be automatically selected in the similar way as the last landmark point of the ASM constructed for cortical plates.

\textbf{V. CONCLUSIONS}

From the work conducted the following conclusions can be drawn:

(i) A fully automated method was developed for the selection of tibial trabecular bone (TB) regions on x-ray images of knee joints. The automated method consists of three components: image preprocessing, delineation of cortical bone plates (active shape
model) and location of regions of interest (ROIs). Thus, time-consuming and prone-to-human-error process associated with the manual selection of bone ROIs can be replaced using the method developed.

(ii) The automated method was evaluated against the "gold standard" containing regions selected by a radiologist expert. The size of each region was 12.8×12.8mm. Similarity index (>-0.8) and offsets [-1.78, 1.27]×[0.26, -0.65]mm (medial) and [-2.15, 1.59]×[0.52, -0.58]mm (lateral) calculated showed that the automated method was accurate for the selection of bone regions.

(iii) These results were confirmed by Bland and Altman plots constructed for FSs and by changes of FDs to region offsets calculated between the "gold standard" and automatically selected regions. The plots showed a random scatter and the 95% confidence intervals of differences were (-0.006, 0.008) and (-0.001, 0.011). The changes of FDs to offsets were not greater than 0.035. This is less than differences in FDs between non-OA and OA bone regions (>-0.05). This indicates that the automated method developed has potential for use in detecting TB changes in OA knee joints.

(iv) ROIs were selected by a second radiologist. They were closer to the "gold standard" than those produced by the automated method. However, the manual selection is expensive and it cannot be guaranteed that follow-up studies will be performed
by the same radiologist or a radiologist with the same experience.

(v) The newly developed method will be used in the development of a fully automated system for prediction, early detection and diagnosis of the knee osteoarthritis. We hope that the automated system can improve the accuracy of predicting and detecting OA so that one can start appropriate therapy. Our current work is focused on the development of automated methods for the analysis of joint space narrowing, osteophytes, leg alignment and gait patterns.

ACKNOWLEDGMENTS

The authors wish to thank Simone Smattfield, Steve Morris from the Perth Radiographic Clinic and Karl Stoffel from Fremantle Hospital for their help in analysing x-ray radiographs. Support from the Australian Research Council, the National Health and Medical Research Council and the Department of Mechanical Engineering, University of Western Australia, for this project, is greatly appreciated. The authors also wish to thank the reviewers for the time they spent reviewing this manuscript and their helpful and constructive comments.


2E.A. Messent, R.J. Ward, C.J. Tonkin, and J.C. Buckland-Wright, “Cancellous bone difference between knees with early, definite and advanced


10E.A. Messent, R.J. Ward, C.J. Tonkin, and J.C. Buckland-Wright, "Differences in trabecular structure between knees with and without osteoarthritis
quantified by macro and standard radiography, respectively." Osteoarthritis and Cartilage 14, 1302-1305 (2006).


TABLE I. Mean, number (n), 95% confidence range (CR) and interval (CI) values of offsets in millimetres and similarity indices (SI) calculated between the "gold standard" regions and bone regions selected by the second radiologist (Manual) and the automated method (Automated). Only cases that were offset in the positive (negative) direction were included in the calculation of the mean and 95% CR and CI values in the positive (negative) direction.

<table>
<thead>
<tr>
<th>Knee compartment</th>
<th>Offset and SI</th>
<th>n</th>
<th>Manual [mm]</th>
<th>Automated [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean</td>
<td>95% confidence range and interval</td>
<td>n</td>
</tr>
<tr>
<td>Medial</td>
<td>-Δx</td>
<td>44</td>
<td>-0.35</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>+Δx</td>
<td>61</td>
<td>0.38</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>+Δy</td>
<td>44</td>
<td>0.44</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>-Δy</td>
<td>69</td>
<td>-0.39</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>SI</td>
<td>132</td>
<td>0.93</td>
<td>0.01</td>
</tr>
<tr>
<td>Lateral</td>
<td>-Δx</td>
<td>55</td>
<td>-0.36</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>+Δx</td>
<td>48</td>
<td>0.36</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>+Δy</td>
<td>51</td>
<td>0.57</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>-Δy</td>
<td>67</td>
<td>-0.49</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>SI</td>
<td>132</td>
<td>0.92</td>
<td>0.01</td>
</tr>
</tbody>
</table>

-Δx (+Δx) - horizontal offset toward the medial (lateral) site of the tibia bone
-Δy (+Δy) - vertical offset toward the distal (proximal) tibia bone

TABLE II. Times required for training the automated region selection method and selecting manually and automatically a new ROI (a novel case).

Computation was performed using Unix computer with 1.2GHz clock.

<table>
<thead>
<tr>
<th>Region selection</th>
<th>Task</th>
<th>Time [min]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>Selection of &quot;gold standard&quot; ROIs (~1min per region)</td>
<td>~80</td>
</tr>
<tr>
<td>(40 knee images; 2 regions per image)</td>
<td>Clicking 30 landmark points on the cortical plate border (~1.5min per region)</td>
<td>~120</td>
</tr>
<tr>
<td></td>
<td>ASM training (~1min per region)</td>
<td>~80</td>
</tr>
<tr>
<td></td>
<td>Clicking 10 points on the fibula head (~0.5min per image)</td>
<td>~20</td>
</tr>
<tr>
<td></td>
<td>Finding optimal thresholds (~271min per image)</td>
<td>~10840</td>
</tr>
<tr>
<td></td>
<td>Total training time</td>
<td>~11140</td>
</tr>
<tr>
<td>Automated selection</td>
<td>Selection of a new ROI</td>
<td>~0.5</td>
</tr>
<tr>
<td>Manual selection</td>
<td>Selection of a new ROI</td>
<td>~1</td>
</tr>
</tbody>
</table>
FIG. 1. ROIs selected manually by a radiologist expert.
FIG. 2. Pre-processing of a knee x-ray image: (a) the original image resized to 512 × 512 pixels; (b) enhanced image obtained after applying a CLAHE method combined with a median filtering and a Sobel edge sharpening to the original image; (c)-(d) knee borders and joint space line found on the enhanced image.
FIG. 3. Delineation of the cortical bone plate: (a) an example of landmark points selected manually by a human observer. The points were used to construct a mean shape model; (b) points (marked as crosses) used to initialize landmark points for the shape model.
FIG. 4. Initial positions of ROIs located immediately under the lowest points of the cortical bone plates.

FIG. 5. Positions of ROIs obtained after applying the horizontal adjustment.
FIG. 6. A vertical adjustment of the ROI position to avoid a subchondral bone sclerosis. A rectangular region (the dotted line region) was constructed by enlarging the ROI (the solid-line region) in the vertical and horizontal direction. The ROI was moved vertically under the white area representing the bone sclerosis found within the enlarged region.
FIG. 7. A final adjustment of the ROI position to avoid the fibula head. An ellipse was fitted to two points representing the superior aspect [the solid dot shown in (a) and (j)] and the medial aspect of the fibula head [the cross shown in (j)]. The center point of the ellipse [the star shown in (j)] is a cross-point of the horizontal and vertical lines running through the medial and superior aspect of the fibula head, respectively. Details of image processing steps shown in (c)-(i) are explained in text. The ROI was moved horizontally to avoid an overlap with the ellipse and subsequently, with the fibula head; as shown in (k).
FIG. 8. A similarity index (SI) measure between the gold standard ROI and the ROI selected by the automated method. TP (true positive) - correctly selected pixels; FN (false negative) - pixels of ROI missed by the automated method; FP (false positive) - falsely selected pixels of ROI as belonging to the gold standard (FPa - pixels on background, FPb - pixels on fibula).
(a) medial compartment

(b) lateral compartment
FIG. 9. Plots of similarity index SI values against thresholds: (a) medial and (b) lateral compartment, and (c) fibula head.
(a) Mean = 0.001; SD = 0.043
95%CI = (-0.006, 0.008)

(b) Mean = 0.004; SD = 0.038
95%CI = (-0.001, 0.011)

(c) Mean = -0.003; SD = 0.039
95%CI = (-0.010, 0.002)

(d) Mean = 0.0001; SD = 0.023
95%CI = (-0.004, 0.004)

FIG. 10. Bland and Altman plots of FDs calculated in the vertical and horizontal direction at scale of 0.65mm for the "gold standard" versus (a,b) the automated method (Automated) and (c,d) the second radiologist (Manual).
FIG. 11. Images of mean values of differences in fractal dimensions calculated in the (a) horizontal and (b) vertical direction at scale of 0.65mm between the 132 "gold standard" regions and their offsets. Mean values (the solid-line rectangle) and 95% CI (the dotted-line rectangles) of offset distances calculated between the "gold standard" and automatically selected regions were superimposed on the images. Results obtained from the lateral site of knee joints were used.
FIG. 12. Changes in FDs calculated in the (a) horizontal and (b) vertical direction at scale of 0.65 mm for a ROI overlapped by fibula head.
FIG. 13. Examples of poorly selected ROIs (white) by the automated method and the gold standard regions (black). Errors occurred because of: (a) knee borders were incorrectly identified and (b)-(c) ellipses underestimated or overestimated the fibula head.