

A CAD System for Long-Bone Segmentation and Fracture Detection

Martin Donnelley, Greg Knowles, and Trevor Hearn

Flinders University, GPO Box 2100, Adelaide 5001, South Australia
{martin.donnelley, greg.knowles, trevor.hearn}@flinders.edu.au

Abstract. Medical imaging has advanced at a tremendous rate since x-rays were discovered in 1895. Today, x-ray machines produce extremely high-quality images for radiologists to interpret. However, the methods of interpretation have only recently begun to be augmented by advances in computer technology. Computer aided diagnosis (CAD) systems that guide healthcare professionals in making the correct diagnosis are slowly becoming more prevalent throughout the medical field. Detection of long-bone fractures is an important orthopaedic and radiologic problem, and it is proposed that a novel CAD system could help reduce the number of fractures missed during x-ray diagnosis. A number of image processing software algorithms useful for assisting the fracture detection process are described, and their accuracy evaluated on a database of fracture images from trauma patients. Incorporating these methods will further expand the capabilities of today's CAD systems, and result in more accurate diagnosis of fractures and a reduction of the fracture miss rate.

Keywords: Bone, CAD, Fractures, AMSS, Hough, Segmentation.

1 Introduction

Fractures of bone are a common affliction, accounting for approximately 20% occupancy of orthopaedic wards at any given time [1]. In the Australian population the number of fractures associated with age-related bone loss is increasing rapidly, with the number of non-hip fractures estimated to increase by 9% every 5 years until 2036 [2].

Trained radiologists generally identify abnormal pathologies including fractures with a relatively high level of accuracy. However studies examining reader accuracy [3] have shown that in some cases the miss rate can be as high as 30%, or even higher when reading x-rays containing multiple abnormalities. Accurate diagnosis of fractures is vital to the effective management of patient injuries. As a result, detection of long-bone fractures is an important orthopaedic and radiologic problem, and it is proposed that a novel CAD system could help lower the miss rate. This paper examines the development of such a system, for the detection of long-bone fractures.

To produce a CAD system for long-bone fracture detection, two major goals were identified. The first was to create a semi-automatic long-bone segmentation

method to separate the diaphysis from the proximal and distal epiphyses. The second was to create a semi-automatic long-bone fracture detection method that could identify fractures within the diaphyseal segment.

2 The Fracture Detection Algorithms

Our CAD system consists of four stages that are described in the following sections. For development and evaluation a set 50 long-bone images obtained from the Flinders Medical Centre Emergency Department was split into 6 training images and 44 testing images. All training images contained at least one fracture, while 38 of the 44 test images contained at least one fracture.

2.1 Long-Bone Edge Detection

Since the structures within an x-ray image were not easily computed, we used a scale-space approach to detect edges. Depending on the scale chosen, local disturbing detail could be removed and the features at that scale could be extracted. To compute features within an image $I(x, y)$, $(x, y) \in R$, scale-space theory proposes that the image be replaced by a smoothed version $I(x, y, t)$, where the degree of image smoothing is determined by the scale parameter t . After smoothing, the points with high gradient intensity $|\nabla I(x, y, t)|$ contain the most significant global information [4].

To meet all invariance requirements—including properties like Euclidean (translation and rotation) invariance—all multi-scale image analysis techniques should have the form of a curvature motion [5]. We were interested in a special case called the affine morphological scale space (AMSS):

$$\frac{\partial I}{\partial t} = |\nabla I| (\text{curv}(I))^{\frac{1}{3}}. \quad (1)$$

This case was chosen because it diffused purely in the direction orthogonal to the gradient, plus it was also the only case that produced an additional invariance termed affine invariance. A curve is said to be affine invariant if the coordinate system in which it is represented can change without affecting the relative geometry of the curve. An example image and the magnitude of the gradient—calculated using the AMSS—are shown in Figure 1.

2.2 Parameter Approximation

Rather than attempting to accurately segment the bone from the background using traditional segmentation algorithms, only the information about the location and orientation of the bone was determined. To extract position information, prior knowledge about expected human long-bone morphology was utilised. Regardless of their anatomical location, long-bone diaphyses containing no major

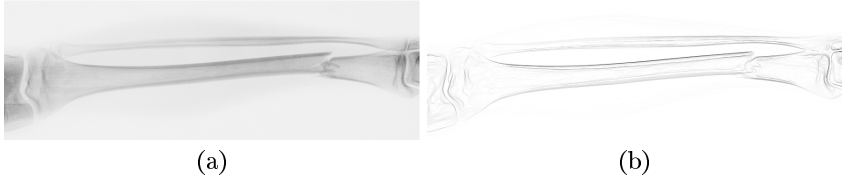


Fig. 1. (a) The original image, and (b) the magnitude of the gradient $|\nabla I(x, y, t)|$ calculated using the AMSS

pathology were typically very straight despite a large degree of natural variation in bone shape. This allowed the diaphyseal bone edges to be approximated by line straight parameters useful for both diaphysis segmentation and fracture detection.

Global information for approximating the long bones within the x-ray image was extracted from the magnitude of the gradient $|\nabla I(x, y, t)|$ using the normal parameterization of the Hough transform. Hence spatially extended patterns were converted into spatially compact features within the space of possible parameter values (ρ, θ) . Thus the problem was reduced from a difficult global feature analysis to a simpler local peak detection in the parameter space.

Peak detection in the resulting Hough accumulator H is a non-trivial exercise, especially with complicated input images. Simple peak detection algorithms that locate the absolute maximum or utilise a global threshold often work well for artificial images, but generally perform poorly when applied to real x-ray images. Even for an infinitesimally thin line, oversampling in the parameter space causes the peak in the accumulator to spread rather than remain localised to one cell. Thus, in practice the votes do not fall into uniform and compact regions, so the peaks are often poorly defined and spread in both the ρ and θ directions. Real world images containing thicker, non-straight lines—such as the edges $|\nabla I(x, y, t)|$ in a typical x-ray image—cause additional spread. The amount of expected spread was quantified using [6], and was used to iteratively locate the maximum elements in the accumulator and thus obtain a set of parameters (ρ_i, θ_i) to approximate the long-bone shaft.

Although the bone edges were approximated by the parameters (ρ_i, θ_i) , no information regarding the length or the endpoints of the lines was known. To accurately detect the long-bone shaft from the peak parameters, an endpoint detector based on [7] was created. In this technique, the path along each parameterised line was examined using a two stage thresholding technique. The first threshold was applied to the path so that only regions where the line matched the bone edge were retained, while the second was applied to ensure that only lines of a minimum length were retained. Figure 2b shows the lines detected by the long-bone shaft endpoint detector. The endpoints were located where the bone edge deviated sufficiently away from the line.

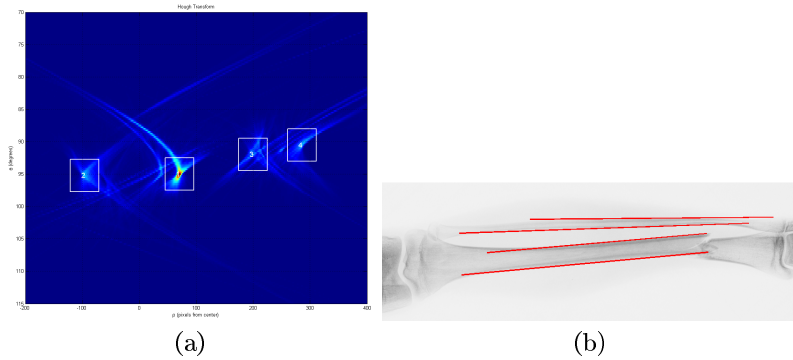


Fig. 2. (a) The Hough Transform of the image in Figure 1b, showing the detected peaks (and the estimate of the peak spread). (b) The lines detected using the long-bone shaft endpoint detector.

2.3 Diaphysis Segmentation

The first segmentation stage used the long-bone approximation parameters (ρ_i, θ_i) to calculate the best estimate of the bone centre-line. In general, centre-line detection was much more difficult for displaced fractures (those in which the centre-lines of the shaft segments on either side of the fracture remained colinear) than undisplaced fractures, and also for distal bones than proximal bones.

In cases where $i > 2$ it was necessary to determine which lines should be paired together. Lines that were to be paired should be located relatively close to each other, because the bone edges should in most cases be approximately parallel. This also meant that when correctly paired, the line endpoints should also be relatively close together. Thus it was proposed that the criterion for determining the correct pairings would be the minimum total distance between the line endpoints. So all the possible pairings were analysed and the combination that minimised this distance was chosen. When multiple centre-lines were detected for the distal long-bones, only the one along the axis of the wider bone was retained. An example of the relationship between the detected bone edges (red) and resulting bone centre-line (green) is shown in Figure 3a.

By definition the boundary between the diaphysis and the epiphysis is the point at which the long-bone shaft begins to widen and curve. Fortunately, the long-bone shaft endpoint detector was used to locate those points at which the bone deviated sufficiently away from the parameterised line fitted to the straight portion of the bone edge. Since the line endpoints on either side of the shaft often varied, it was appropriate to use an average measure, which was the same point at which the centre-line ended. The segmentation boundary at the diaphysis should also be orthogonal to the centre-line at that point. Therefore at both ends of every identified centre-line, an orthogonal line spanning the image was added, as shown dashed in blue in Figure 3a. Regardless of the number of centre-lines in the image, it was always the two orthogonal lines furthest apart that

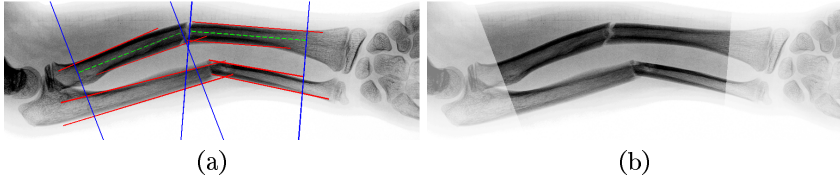


Fig. 3. (a) Once the centre-lines (*green dashed*) had been located from all of the bone edges (*red solid*), lines orthogonal to the centre-line endpoints (*blue solid*) were added. From these lines the segmentation was performed. (b) The segmentation result.

corresponded to the diaphyses. An example diaphysis segmentation is shown in Figure 3b.

2.4 Fracture Detection Using Gradient Analysis

Long bone diaphyses consist almost exclusively of uniform cortical bone. X-rays of a typical diaphysis containing no abnormal pathology should exhibit gradient changes in the direction normal to the bone axis, due to the bone-tissue interface, and the variation in cortical wall thickness through which the x-rays must pass. The wall thickness and density change relatively slowly along the length of the long bone, so the gradient parallel to the bone edges is small. Large gradients occurring at angles not orthogonal to the bone edges are therefore indicative of some type of abnormality, and are detected by this algorithm.

In 2.2 line parameters to approximate the edges of the long bones were determined, and from this the magnitude and direction of the edges that should be present within the long-bone shaft are known. Removing these leaves behind only the abnormal gradients, such as those that belong to any fractures. To remove all normal regions from the image a composite measure of the magnitude $|\nabla I(x, y, t)|$ and direction $\phi(x, y, t)$ of the gradient of the smoothed image $I(x, y, t)$ at scale t , where t is suitably small for microscopic feature analysis is used.

The gradient composite measure (GCM) $C(x, y, t, \rho, \theta, p)$ is the product of the magnitude of the gradient and two scaling factors used to incorporate the direction of the gradient $\phi(x, y, t)$, with the long-bone shaft approximation parameters. The first scaling factor—called the importance rank $R(x, y, t, \theta, p)$ —measured how well the direction of the gradient matched the angle θ of the approximation lines, while the second—called the distance rank $D(x, y, \rho, \theta, p)$ —measured how close each pixel was to all the approximation lines:

$$C(x, y, t, \rho, \theta, p) = |\nabla I(x, y, t)| R(x, y, t, \theta, p) D(x, y, \rho, \theta, p) . \quad (2)$$

The magnitude data was normalised to the range $[0, 1]$. However, the direction data was in the range $[0, 360^\circ]$ and had to be ranked in terms of which angle was most important, since large angles did not necessarily correspond to more

important regions. The importance rank $R(x, y, t, \alpha, p)$ was based on the direction of the gradient $\phi(x, y, t)$ at that pixel, a chosen transform angle α , and an importance weighting coefficient p . It was calculated such that the transform angles α and $\alpha \pm 180^\circ$ were assigned the minimum value 0, while orthogonal angles $\alpha \pm 90^\circ$ were assigned the maximum value 1. Fractures and other abnormalities were located by setting the transform angle α equal to the angle parameter θ_i of each of the approximating lines:

$$R(x, y, t, \theta, p) = \prod_{i=1}^e \left(1 - [\cos(\phi(x, y, t) - \theta_i)]^{2p} \right) . \quad (3)$$

Similarly, the distance rank was a measure of how far a point in the image was from all of the long-bone shaft approximation lines. For a single line with parameters (ρ, θ) the distance rank for a point (x, y) was simply the length of the normal between that point and the approximation line. A power relationship p was again applied, so that regions close to the line were given a much higher importance than those further from the line. The distance rank was calculated for each point in the image using:

$$D(x, y, \rho, \theta, p) = \prod_{i=1}^e \left(|\rho_i - [(x - x_{origin}) \cos \theta_i + (y_{origin} - y) \sin \theta_i]|^{\frac{1}{p}} \right) . \quad (4)$$

As shown in Figure 4a, the GCM did indeed indicate the location of the fracture, but some further processing was used to make the diagnosis clearer. Firstly a 9 x 9 median filter was used to remove salt and pepper type noise created by random matches between ϕ and θ in the GCM calculation. The regions most likely to correspond to a fracture were isolated by simultaneously applying two empirically chosen thresholds, only keeping the pixels that had a sufficiently large composite measure (greater than T_1), and also occurred in a large enough cluster to be significant (sum of the pixel intensities in the cluster was greater than T_2). Once these regions were identified they were enlarged to a user adjustable size using morphological dilation and their boundaries were marked on the original image. Figure 4b demonstrates the result.

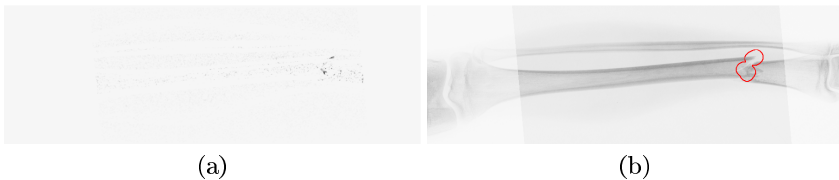


Fig. 4. (a) The gradient composite measure (b) Clustering was applied to mark the identified regions

3 Evaluation

3.1 Long-Bone Edge Detection

Applying the algorithm to the testing images showed that the smoothing produced by the AMSS algorithm was contained within the boundaries of the bone, and that increasing the scale definitely resulted in greater smoothing.

3.2 Parameter Approximation

When the Hough Transform and peak detection methods were applied a total of 160 peaks were detected. Manual inspection of these peaks showed that 142 (88.8%) were in locations that correctly corresponded to the bone shaft edges. The incorrect peaks were dispersed across the 44 images such that only 12 images (27.3%) were completely correct (no false peaks). The highest rates of detection occurred in the humerus and femur, where all the detected peaks corresponded to points of interest on the shaft. Once all 185 peaks were correctly located, the line endpoint detector was applied. Manual inspection showed that only four of the 185 lines (2.2%) were incorrect. The incorrect lines were in four different images from the 44 image set, corresponding to 9.1% incorrect images. The incorrect lines all occurred as a result of good matches with edges belonging to other features within the image that happened to lie underneath the line of interest.

3.3 Diaphysis Segmentation

When tested, the centre-line detection algorithm correctly located 52 of the 56 centre-lines (93%) manually identified in the 44 images. In all four cases where a centre-line was not detected correctly, the cause was that the lines were not appropriately paired. In these cases the incorrect detection of the centre-line had little effect on the final segmentation, although this would not always be the case. Of the 71 segmentation boundaries in the test set, 59 were identified correctly, corresponding to a detection rate of 83%. In addition, in 33 out of the 44 images (75%), all the required segmentation boundaries were correctly detected, such that those images were segmented completely correctly. The most common cause of choosing the incorrect location for segmentation (6 cases) was the fracture disturbing the centre-line such that it ended prematurely, rather than continuing to the point where it would if the fracture were not present. Thus, segmentation was better in images where the bones were complete—that is, images containing subtle fractures with no displacement. In fact, when a subset of 24 images containing only subtle undisplaced fractures was used—rather than the complete set that also contained obvious grossly displaced fractures—the diaphysis segmentation rate was 100%.

3.4 Fracture Detection Using Gradient Analysis

Using a sensitivity analysis, values of $p = 5$, $T_1 = 0.008$ and $T_2 = 50$ were chosen. A manual comparison to the radiologist's results was performed, and the

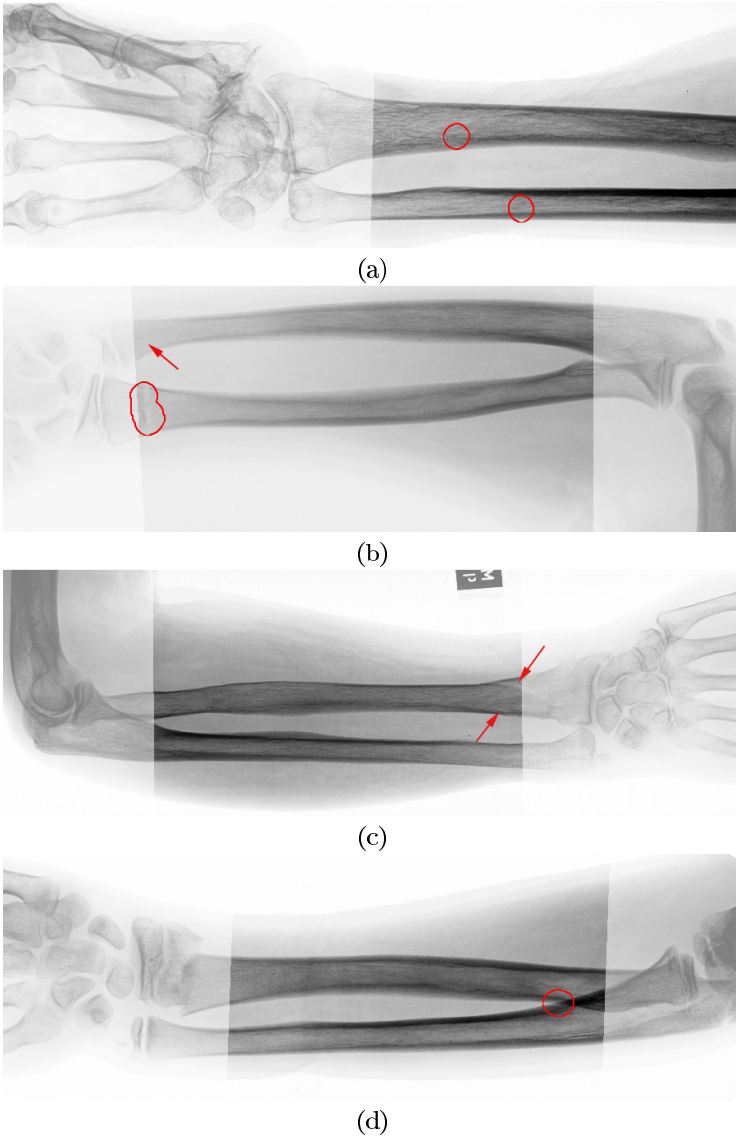


Fig. 5. Examples of the results produced by the fracture detection algorithm. (a) Two extremely subtle midshaft forearm fractures were detected correctly. (b) One subtle radius fracture was detected correctly, but the corresponding subtle ulna fracture (arrow) was missed. (c) Subtle radius and ulna fractures (arrows) were missed. (d) A false detection due to overlapping bones.

Table 1. (a) A confusion matrix showing the results produced by the algorithm on a sample of 47 fractures present in the 44 image test set. (b) The causes of false detection.

	Predicted Non-fracture	Predicted Fracture
Non-fracture	1 (11%)	68
Fracture	8 (17%)	39 (83%)

(a)

	Cause of false detection	Number	Individual %	Group %
Algorithm	Bone overlap	7	10.3%	27.9%
	Image artifact	12	17.6%	
Biological	Harris growth arrest line	5	7.3%	72.1%
	Feature not related to fracture	38	55.9%	
	Soft tissue injury	6	8.8%	

(b)

number of true and false positive and negative results were recorded, along with the causes of any false positives. The test image set contained a total of 47 fractures, 39 of which were correctly detected by the algorithm, corresponding to a detection rate of 83% of all fractures. A summary of the results is shown in the confusion matrix in Table 1a. The causes of all the false positives were recorded, and were split into two categories: algorithm errors, and misinterpreted biological phenomena. Details are shown in Table 1b. Example results produced by the algorithm are shown in Figure 5.

4 Conclusion

Computer aided detection of midshaft long-bone fractures has not previously been examined. A method by which semi-automated long-bone shaft segmentation, along with fracture detection within the segmented region has been presented. In the test set, 83% of the diaphysis segmentation boundaries were correctly identified, and 83% of the fractures within those segmented regions were also detected correctly. These results will further expand the capabilities of today's CAD systems, and result in more accurate diagnosis of fractures and a reduction of the fracture miss rate.

References

1. Chipchase, L.S., McCaul, K., Hearn, T.C.: Hip fracture rates in South Australia: Into the next century. *Australian and New Zealand Journal of Surgery* 70(2), 117–119 (2000)
2. Sanders, K.M., Nicholson, G.C., Ugoni, A.M., Pasco, J.A., Seeman, E., Kotowicz, M.A.: Health burden of hip and other fractures in Australia beyond 2000. Projections based on the Geelong Osteoporosis Study. *Medical Journal of Australia* 170(10), 467–470 (1999)
3. Berlin, L.: Reporting the missed radiologic diagnosis: Medicolegal and ethical considerations. *Radiology* 192(1), 183–187 (1994)

4. Marr, D.: Vision: A computational investigation into the Human Representation and Processing of Visual Information. W. H. Freeman and Company, New York (1982)
5. Alvarez, L., Lions, P.L., Morel, J.M.: Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal of Numerical Analysis* 29(3), 845–866 (1992)
6. Atiquzzaman, M.: Multiresolution Hough transform - An efficient method of detecting pattern in images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(11), 1090–1095 (1992)
7. Skingley, J., Rye, A.: The Hough transform applied to SAR images for thin line detection. *Pattern Recognition Letters* 6(1), 61–67 (1987)