DETECTION OF FEMUR AND RADIUS FRACTURES IN X-RAY IMAGES

Sher Ee Lim[1], Yage Xing[1], Ying Chen[1], Wee Kheng Leow[1], Tet Sen Howe[2], and Meng Ai Png[3]

[1]Dept. of Computer Science, National University of Singapore
3 Science Drive 2, Singapore 117543
[2]Dept. of Orthopaedics, Singapore General Hospital
Outram Road, Singapore 169608
[3]Dept. of Diagnostic Radiology, Singapore General Hospital
Outram Road, Singapore 169608

limshere, xingyage, chenying, leowwk@comp.nus.edu.sg, tshowe@sgh.com.sg

Abstract: 13% of men and 30%–40% of women suffer from osteoporotic bone fractures worldwide. In large hospitals, doctors need to visually inspect a large number of x-ray images to identify the fracture cases, which typically constitute a small fraction of all the x-ray images examined. After looking through many images containing healthy bones, a tired radiologist has been found to miss a fractured case among the many healthy ones. Automated fracture detection can help the doctors by screening for obvious cases and flagging suspicious cases for closer examinations. Since bone fractures can occur in many ways, no one single algorithm can detect all the possible fractures accurately. This paper describes an approach in detecting fractures of the femur and the radius by combining various detection methods. These methods extract different kinds of features for fracture detection. They include neck-shaft angle, which is specifically extracted for femur fracture detection, and Gabor texture, Markov Random Field texture, and intensity gradient, which are general features that can be applied to detecting fractures of various bones. Two types of classifiers are tested, namely, Bayesian classifier and Support Vector Machine. Test results show that the combined approach can improve both the fracture detection rate and the classification accuracy significantly compared to any single method.

Keywords: image analysis, image interpretation and understanding, x-ray fracture detection, femur fractures, radius fractures.

INTRODUCTION

Many people suffer from bone fractures worldwide. The International Osteoporosis Foundation [10] reported that, worldwide, women has a 30%–40% lifetime risk of getting osteoporotic fractures while men has a lower risk of 13%. The number of hip fractures could rise from 1.7 million worldwide in 1990 to 6.3 million by 2050. The most dramatic increase is expected to occur in Asia during the next decades. World Health Organization confirms that osteoporosis is second only to cardiovascular disease as a leading health care problem [10].

In clinical practice, doctors and radiologists in large hospitals have to visually inspect x-ray images to determine the occurrence and the precise nature of fractures. Typically, only a small fraction of the images contain fractures. For example, in the x-ray images that we collected from a local public hospital, only about 12% of the femurs are fractured. After looking through many images containing healthy bones, a tired radiologist has been found to miss a fractured case among the many healthy ones. A computer vision system can assist the doctors by screening the x-ray images for obvious cases and flagging suspicious cases for closer examinations. Automated screening of x-ray images can thus help to reduce the radiologists' workload and direct the doctors' attention to suspicious cases to improve the timeliness and accuracy of their diagnosis. Therefore, such a computer vision system is extremely useful for clinicians and is now feasible because all clinical radiology is going digital. Digital x-ray images are now routinely captured using digital x-ray machines.

Fractures of a bone can occur in many ways. As such, no one single method can accurately detect all kinds of bone fractures. Our approach is to develop a suite of methods, each designed to complement the others by looking for different features that characterize different kinds of fractures. This paper presents our current research work on the detection of bone fractures. We will focus mainly on the detection of femur fractures since they are the most common type of fractures. Some preliminary results on the detection of fractures of the radius near the wrist will also be reported.

RELATED WORK

Our first published work on the detection of fractures in x-ray images is that of Tian et al. [22]. The method detects femur fractures by computing the angle between the neck axis and shaft axis. Apparently, this method can only detect severe fractures that cause significant changes in the neck-shaft angles (Fig. 1(a)). Some fractures may crack the neck without rotating or displacing the femoral head significantly. They result in only local disruptions of the trabecular patterns (Fig. 1(b)).

Following the first attempt, Yap et al. [26] developed a complementary method for detecting femur fractures by analyzing the disruption of trabecular pattern at the femoral neck. This was done by extracting and analyz-

(a) (b)

Fig. 1: (a) Severe fracture that changes the shape of the femur. (b) Fracture that does not change the femur's shape significantly.

ing the orientations of the trabeculae using Gabor filters [1]. Test results showed that this method complemented the neck-shaft angle method. Hence, combing the two methods improved the performance of fracture detection.

Other related methods used non-visual techniques to detect fractures. For example, Ryder et al. analyzed acoustic pulses to determine whether a fracture had occurred [19]. Kaufman et al. analyzed mechanical vibration using a neural network model [11] whereas Singh and Chauhan measured electrical conductivity [21].

Most research effort on the analysis of orthopaedic x-ray images has been focused on the detection of osteoporosis by methods of texture analysis and fractal analysis. For example, [8, 13, 18] computed first order statistics of texture including mean and standard deviation, and [17, 25] computed second order statistics of texture such as the co-occurrence matrix. On the other hand, [2, 3, 12] applied techniques such as semi-variance, surface area, and power spectral methods to derive the fractal dimension and geometry for detecting osteoporosis. Caligiuri et al. [3] showed that their method appeared promising in discriminating between healthy and fractured bones. [16] applied a multiresolution wavelet technique on the micro x-ray CT images of rats' lumbar vertebrae, and [14] applied fractal analysis on the micro x-ray images of human knees. Both these works analyzed trabecular structures of bones in high-resolution micro x-ray images instead of normal x-ray images. A good review of the application of fractal analysis for the assessment of trabecular bone structure is given in [15].

There is a substantial amount of work on the analysis of tubular structures such as blood vessels and lung bronchi. The analysis of these small structures typically assume certain relationship between image intensity and the position on the structure. For example, the cores method [7] and the ridge detection method [9] have been applied to 2D images to find intensity ridges which correspond to the medial lines of vessels. However, the femur and radius are large structures with complex internal structure of trabeculae (Figs. 1, 7), which show up as complex texture patterns in x-ray images. So, standard methods for analyzing tubular structures cannot work on femur and radius x-ray images.

FRACTURE DETECTION

Our method of detecting fractures can be divided into 3 stages: (1) extraction of approximate contour of the bone of interest, (2) extraction of features from the x-ray image, and (3) classification of the bone based on the extracted features. The extraction of bone contour in stage 1 is performed using active shape model, supplemented by active appearance models [4] at distinct feature points. These are well known methods that have already been reported in the literature. So, this paper will focus on stages 2 and 3.

Four different types of image features are extracted for fracture detection: (1) femoral neck-shaft angle, (2) Gabor filters [1], (3) Markov Random Field (MRF) [6], and (4) intensity gradient. The first feature is specifically extracted for detecting the distortion of shape due to severe femur fracture. The other features can potentially by applied to detecting fractures of various bones. So far, we have applied all of them to detecting femur fractures and MRF to detecting radius fractures.

Computation of neck-shaft angle has been reported in [22]. Here, we shall give a summary of the method. Given the contour of the femur, it computes lines that are normal to both sides of the shaft contour, which we called the *level lines*. The line that passes through the mid points of the shaft level lines gives a good approximation of the shaft axis. The level lines at the femoral head and neck are clustered into bundles. The mean direction of the level lines in the largest bundle that contains long level lines gives an initial approximation of the neck axis. Then, taking note of the symmetry of the femoral head, the best axis of symmetry is determined, starting with the initial approximation, to obtain the best approximation of the shaft axis. Finally, the angle between the neck and the shaft axes is computed. Classification is based on a threshold of the neck-shaft angle that is learned from training samples.

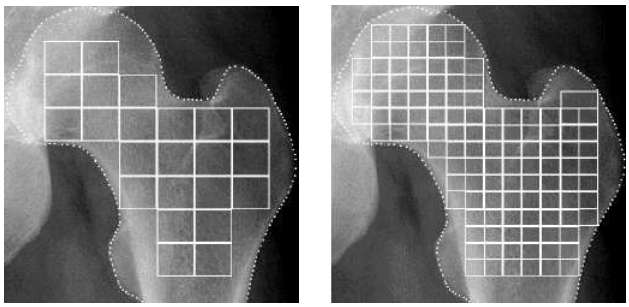
The extraction of the other three features share a common theme: *adaptive sampling*. The shapes and sizes of the bones are not identical in the x-ray images. Even among healthy bones, there are still differences in the appearance because they are naturally-occurring objects. Age and gender also contribute to the difference in the appearance of the bones. One standard method of dealing with size variation is to normalize the size of the bones in the images. This method is, however, unsatisfactory because it can either remove important texture information (if the image is shrunken) or introduce noise and artifacts (if the image is enlarged). Instead of scaling the x-ray images, adaptive sampling is used to sample the features so that the sampled locations in different images correspond to consistent locations in a normalized sampling grid.

Adaptive Sampling

Let W and H denote the width and height of the bounding box that contains the bone of interest, e.g., the femur's upper extremity. This bounding box is automat-

Table 1: Number of sampling locations for various feature types. MRF: Markov Random Field model, IG: intensity gradient. Gabor and IG are extracted only from femur images.

	Gabor	IG	MRF (femur)	MRF (radius)
n_x	12	28	16	8
n_y	14	32	24	15



(a) (b)

Fig. 2: Adaptive sampling grid. (a) Without overlap. (b) With overlaps.

ically derived from the approximate bone contour extracted in Stage 1 of the algorithm. Let n_x and n_y denote the number of sampling locations along the x- and y-directions, which means the sampling method divides the whole bounding box into $n_x \times n_y$ regions. To improve the coverage of the grids on the image, the sampling grids are arranged in such a way that neighboring regions overlap by 50% of their width and height. So each region is of width S_x and height S_y given by:

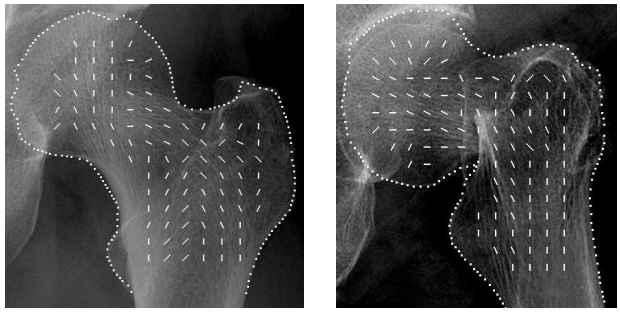
$$S_x = 2\left\lfloor \frac{W}{n_x+1} \right\rfloor, \quad S_y = 2\left\lfloor \frac{H}{n_y+1} \right\rfloor. \qquad (1)$$

The advantage of this adaptive sampling method is that it requires the extraction of only approximate bone contours. Therefore, it can also tolerate slight variation of shape over different patients, and does not require very accurate extraction of the bone contours.

The number of sampling locations differ for different feature types. Texture features extracted using Gabor filtering requires a larger sampling region and thus fewer sampling locations. On the other hand, extraction of intensity gradient requires very small sampling region. Thus, sampling can be performed at more locations. Markov Random Field (MRF) texture model extracts features from moderate-sized sampling regions. In the current implementation, the number of sampling locations is set as shown in Table 1. Figure 2 illustrates an example of adaptive sampling at the femoral neck.

Feature Extraction

The trabeculae in the bones are oriented at specific orientations to support the forces acting on them. A fracture of the bone causes a disruption of the trabecular pattern, which can be detected by extracting various feature types. In our implementation, features are extracted from each sampling region determined using adaptive sampling method. Extraction of Gabor textures has already been reported in [26]. The result is a *Gabor texture orientation map* $\mathbf{M}_G = [\mathbf{u}_{ij}]$ where $\mathbf{u}_{ij}$ is a unit vector that represents the Gabor texture orientation at grid location (i, j). Figure 3 illustrates examples of Gabor texture orientation maps. In this paper, we shall focus on the extraction of intensity gradient direction and Markov Random Field features.



(a) (b)

Fig. 3: Gabor texture orientation maps of (a) healthy femur and (b) fractured femur. The short lines indicate trabecular orientations.

Before extracting intensity gradient feature, the x-ray images are first normalized so that their mean intensities and contrasts are similar. A general method of computing the intensity gradient at a point $\mathbf{p}$ is to fit a curve surface on the intensity profile at and around $\mathbf{p}$. Then, the intensity gradient can be computed from analytical geometry. However, this approach is too computationally expensive for our application. Instead, an approximation method is applied as follows. Given a region $R(\mathbf{p})$ centered at $\mathbf{p}$, search within the region for a point $\mathbf{q}$ whose intensity difference d_m is the largest:

$$d_m(\mathbf{p}) = \max_{\mathbf{q}' \in R(\mathbf{p})} |I(\mathbf{p}) - I(\mathbf{q}')|. \qquad (2)$$

Then, the intensity gradient direction $\mathbf{g}(\mathbf{p})$ is computed as the vector difference

$$\mathbf{g}(\mathbf{p}) = \text{sgn}(I(\mathbf{p}) - I(\mathbf{q}))\frac{\mathbf{q}-\mathbf{p}}{d_m(\mathbf{p})} \qquad (3)$$

where sgn(.) is the sign function. So, the direction of $\mathbf{g}$ is defined to point from higher intensity location to lower intensity location (Fig. 4). Intensity gradient direction is computed at each location (i, j) within the bone contour to form the *intensity gradient direction map* $\mathbf{M}_g = [\mathbf{u}_{ij}]$. Gradient direction outside the contour is defined to be the null vector. Figure 5 illustrates examples of intensity gradient direction maps.

Intensity normalization is also performed before extracting Markov Random Field texture. Markov Random Field texture model describes the intensity of a pixel $\mathbf{p}$ as a linear combination of those of its neighbors $\mathbf{q}$ [6]:

$$I(\mathbf{p}) = \sum_{\mathbf{q} \in R(\mathbf{p})} (\theta(\mathbf{p}, \mathbf{q})I(\mathbf{p}+\mathbf{q}) + \epsilon(\mathbf{q})) \qquad (4)$$

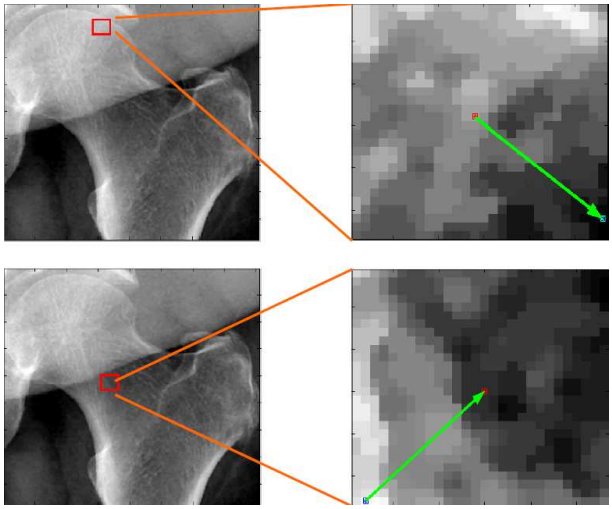
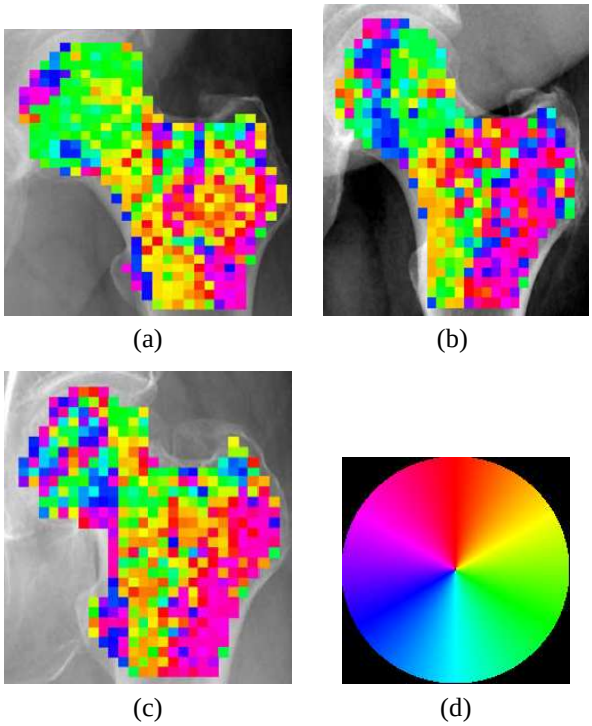


Fig. 4: Intensity gradient directions at different locations.



(a) (b) (c) (d)

Fig. 5: Intensity gradient direction maps of (a, b) healthy femurs and (c) fractured femur. The colors indicate the directions as depicted in (d) the color circle.

where $\theta(\mathbf{p}, \mathbf{q})$ are model parameters and $\epsilon(\mathbf{q})$ represents noise, which is usually assumed to be Gaussian noise of zero mean and constant variance. The model parameters $\theta(\mathbf{p}, \mathbf{q})$ at location $\mathbf{p}$ is then computed by minimizing the error E:

$$E = \left[I(\mathbf{p}) - \sum_{\mathbf{q} \in R(\mathbf{p})} (\theta(\mathbf{p}, \mathbf{q}) I(\mathbf{p} + \mathbf{q}) + \epsilon(\mathbf{q})) \right]^2 . \tag{5}$$

The model parameters $\theta(\mathbf{p}, \mathbf{q})$ are then normalized to unit vectors $\mathbf{u}_{ij}$ to form the MRF texture map $\mathbf{M}_m = [\mathbf{u}_{ij}]$, where $\mathbf{p} = (i, j)$. As for the other feature maps, entries outside the bone contour are assigned the null vectors.

Map Processing

The feature maps described in the previous sections are vector maps, which are not convenient for classification. So they are converted into a *difference maps*, which are scalar maps. For each feature type, the mean feature map of all the healthy training samples, $\overline{\mathbf{M}} = [\mathbf{m_{ij}}]$, is first computed. The entry $\mathbf{m_{ij}}$ is the mean feature vector at position (i, j) in $\overline{\mathbf{M}}$ and it is given by:

$$\mathbf{m}_{ij} = \begin{cases} \sum_{k=1}^{n} \mathbf{u}_{kij} \left\| \sum_{k=1}^{n} \mathbf{u}_{kij} \right\|^{-1} & \text{if } c_{ij} > n/2 \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{6}$$

where n is the number of training samples, $\mathbf{u}_{kij}$ is the unit feature vector of sample k at position (i, j), and c_{ij} is the number of samples with non-null feature vectors at position (i, j).

For a particular position (i, j), if more than half of the training samples' feature maps have null values at this position, it will be considered as an insignificant position, which means this position does not contain significant information for classification. Then, the corresponding entry in the mean feature map will be assigned the null vector $\mathbf{0}$. This situation usually occurs near the boundary contour of the bone because of slight shape variation among different images. By setting the map entries at these positions to $\mathbf{0}$, the effect of slight shape variation on classification can be removed.

Now, the difference map $\mathbf{V} = [v_{ij}]$ for all the training samples can be computed. Each entry v_{ij} indicates the difference between the image's feature map and the mean feature map at position (i, j) and it is given by:

$$v_{ij} = \begin{cases} 0 & \text{if } \mathbf{u}_{ij} = \mathbf{m}_{ij} = \mathbf{0} \\ 1 - |\mathbf{u}_{ij} \cdot \mathbf{m}_{ij}| & \text{otherwise.} \end{cases} \tag{7}$$

A v_{ij} value that is close to 0 indicates a slight difference, and a large v_{ij} indicates a large difference. As the mean feature map is computed over all the healthy training samples, an image of a healthy bone should yield a feature map that is very similar to the mean feature map. So this image's difference map is expected to have mostly

small values. On the other hand, in an image of a fractured bone, there will be some disruption of the trabecular pattern caused by the fracture. So its feature map will be very different from the mean feature map at some positions and its difference map is expected to have some large values. Our classification of the bones will be based on the difference map.

Classification

Two classifiers are applied on the difference maps to classify the test samples: Bayesian and Support Vector Machine (SVM).

For the first method, the sets of healthy and fracture training samples' difference maps are each modeled by a multivariate Gaussian function, which are used to estimate the conditional probabilities $P(x|\,\text{healthy})$ and $P(x|\,\text{fracture})$, where x denotes a sample's difference map. We found that, in practice, the *a priori* probabilities $P(\text{healthy})$ and $P(\text{fracture})$ are not identical. For example, $P(\text{healthy})$ is roughly 0.88 and $P(\text{fracture})$ is 0.12 for the femur images in our collection. Applying Bayes' rule, we obtain

$$P(\text{class}|x) = \frac{P(x|\,\text{class})\,P(\text{class})}{P(x)} \tag{8}$$

where class is either healthy or fractured. The denominator $P(x)$ is the same for both $P(\text{healthy}|x)$ and $P(\text{fracture}|x)$ and so can be ignored. Thus, sample x can be classified as fractured if $P(\text{healthy}|x)$ is smaller than $P(\text{fracture}|x)$.

It is noted that the fractured samples are most likely not well clustered in the feature space. Ideally, one would cluster them into various clusters based on their similarity, and then model each cluster using a different multivariate Gaussian. However, we find that this approach is not appropriate because each cluster contains too few samples for accurate estimation of the multivariate Gaussian.

For the second method, Support Vector Machine [5, 23, 24] is used for classification. The objective of SVM can be stated succinctly as follows:

> Given the training set $\{(\mathbf{x}_i, d_i)\}_{i=1}^{n}$, where d_i is the class of feature vector $\mathbf{x}_i$, find the optimal hyperplane, in terms of weights $\mathbf{w}$ and bias b, that satisfies
>
> $$d_i\,(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \quad \text{for } i = 1, \ldots, n \tag{9}$$
>
> and minimizes $\Phi(\mathbf{w}) = \mathbf{w}^T\mathbf{w}/2$.

The optimal weights $\mathbf{w}$ are given by a set of Lagrange multipliers α_i:

$$\mathbf{w} = \sum_i \alpha_i d_i \mathbf{x}_i\,. \tag{10}$$

The training vectors $\mathbf{x}_i$ with non-zero α_i are the support vectors.

For practical applications, which are typically nonlinearly separable, it is preferred to solve the classification problem in a high-dimensional space where there is a better chance of achieving linear separation. This is accomplished by applying a nonlinear function $\boldsymbol{\phi}(\mathbf{x})$ to map the vector $\mathbf{x}$ in an n-dimensional input space to an m-dimensional feature space, $m > n$. Then, the optimal hyperplane in the m-dimensional feature space is given by:

$$\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}) + b = 0\,. \tag{11}$$

The nonlinear function $\boldsymbol{\phi}(\mathbf{x})$ is a kernel function of the form $K_i(\mathbf{x}) = K(\mathbf{x}, \mathbf{x}_i)$ where $\mathbf{x}_i$ are the support vectors. Then, the optimal hyperplane in the feature space becomes:

$$\sum_i \alpha_i d_i K(\mathbf{x}, \mathbf{x}_i) + b = 0 \tag{12}$$

and the decision function $f(\mathbf{x})$ becomes

$$f(\mathbf{x}) = \sum_i \alpha_i d_i K(\mathbf{x}, \mathbf{x}_i) + b\,. \tag{13}$$

For efficient computation, the kernel functions must satisfy the so-called *Mercer's Theorem* [23, 24]. These kernel functions include:

1. polynomial:

   $$K(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x}^T\mathbf{x}_i + 1)^p \tag{14}$$

   where p is a constant.

2. Gaussian or Radial Basis Function:

   $$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{n\sigma^2}\right) \tag{15}$$

   where σ is the standard deviation of the Gaussian and n is the number of training samples.

3. hyperbolic tangent:

   $$K(\mathbf{x}, \mathbf{x}_i) = \tanh(\beta_0\mathbf{x}^T\mathbf{x}_i + \beta_1) \tag{16}$$

   where β_0 and β_1 are constants. Note that Mercer's theorem is satisfied only for some values of β_0 and β_1.

In the tests, the SVM Toolbox for Matlab implemented by Schwaighofer [20] was applied to perform the classification. The Radial Basis Function (RBF) with parameter $\sigma = 2$ was chosen as the kernel function because RBF kernel was found to yield better classification results. The SVM toolbox allows different weighting factors to be assigned to the errors associated with the training samples. This is useful for encoding different significance levels of the training samples, and for handling imbalanced number of positive and negative training samples. The weighting factor of the healthy femurs was set as 15, whereas that of the fractured femurs was 100. The weighting factor of the fractured femurs was set to a higher value because there were much fewer fractured samples than healthy samples. Empirical test verified that this approach allowed SVM to classify the samples more accurately. The RBF parameter and weighting factors were determined by experimentation on the training samples.

EXPERIMENTS AND DISCUSSION

Femur Fracture Detection

432 femur images were obtained from a local public hospital, and were divided randomly into 324 training and 108 testing images. The percentage of fractured images in the training and testing sets were kept approximately the same (12%). In the training set, 39 femurs were fractured, and in the testing set, 13 were fractured.

Six different classifiers were trained: neck-shaft angle with thresholding, Gabor texture with Bayesian classifier and SVM, intensity gradient direction with Bayesian classifier and SVM, and Markov Random Field texture with SVM. After training, they were applied on the testing samples and three performance measures were computed:

- fracture detection rate:
  the number of correctly detected fractured samples over the number of fractured samples,
- false alarm rate:
  the number of wrongly classified healthy samples over the number of healthy samples,
- classification accuracy:
  the number of correctly classified samples over the total number of samples.

Table 2 shows that individual classifiers have rather low fracture detection rate, particularly IGD with Bayesian classifier and MRF with SVM. However, each of them can detect some fractures that are missed by the other classifiers. So, by combining the classifiers, both the fracture detection rate and classification accuracy can be improved significantly. We have experimented with many different combinations of classifiers and find the following combinations to yield good performance (Table 2):

- 1 of 5:
  A femur is classified as fractured if any one of the five classifiers, except MRF with SVM, classifies it as fractured.
- 1 of 6:
  A femur is classified as fractured if any one of the six classifiers classifies it as fractured.
- 2 of 6:
  A femur is classified as fractured if any two of the six classifiers classify it as fractured.
- 2 of 4:
  A femur is classified as fractured if any two of the following four classifiers classify it as fractured: neck-shaft angle method, Gabor texture with Bayesian classifier, Gabor texture with SVM, and intensity gradient direction with SVM.

The 1-of-5 method has the highest fracture detection rate of 100%, which means every fracture can be detected by at least one of the classifiers. These detected fractures include very subtle fractures (Fig. 6). The test results show that these classifiers can indeed complement each other. The 1-of-6 method also has a fracture detection rate of 100% but a slightly higher false alarm rate of 11.4%, resulting in a slightly lower overall classification accuracy of 88.9%. This is due to the lower classification accuracy of MRF with SVM compared to most of the other methods. The 2-of-6 method has the best overall performance of high fracture detection rate (92.2%), low false alarm rate (1.0%), and high classification accuracy (98.2%). The 2-of-4 method has no false alarm at all, at the expense of lower fracture detection rate (76.9%) and slightly lower classification accuracy (97.2%)

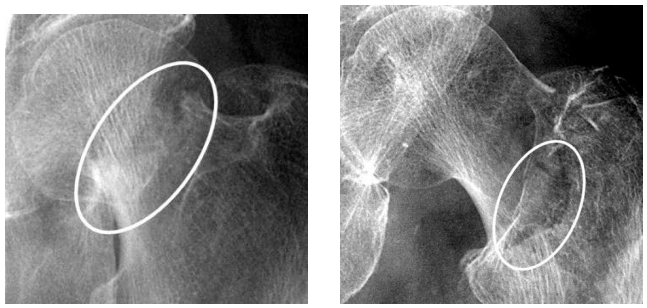


Fig. 6: Subtle fractures at femoral neck (in white ellipse).

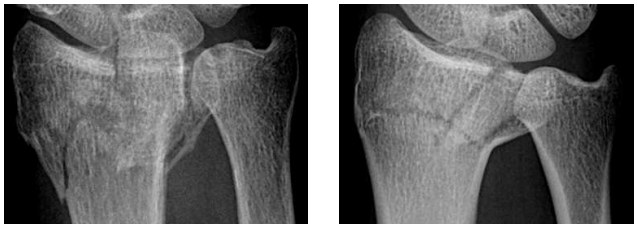


Fig. 7: Sample radius fractures near the wrist.

Radius Fracture Detection

A preliminary test on detecting fractures of the radius near the wrist was also performed using MRF texture model together with SVM classifier. 145 wrist images were obtained from the same local public hospital, and were divided randomly into 71 training and 74 testing images. In the training set, 21 radius bones were fractured whereas 23 were fractured in the testing set. Figure 7 shows examples of radius fractures.

Table 3 shows the performance of the classifier on the testing samples. Interestingly, MRF with SVM performed quite well in detecting radius fractures although it did not perform as well in detecting femur fractures. The reason could be that the fractures of the radius near the wrist are visually more obvious than those at the femoral neck, which can be very subtle (e.g., Fig. 6). So, we expect other feature-classifier combinations to be able to complement MRF with SVM for detecting radius fractures as well.

Table 2: Comparison of performance on femur fracture detection. NSA: neck-shaft angle method, IGD: intensity gradient direction, MRF: Markov Random Field.

		Gabor		IGD		MRF	Combined			
	NSA	Bayes	SVM	Bayes	SVM	SVM	1 of 5	1 of 6	2 of 6	2 of 4
fracture detection rate	61.5%	53.9%	69.2%	38.5%	53.9%	14.3%	100%	100%	92.2%	76.9%
false alarm rate	1.9%	0.0%	2.9%	4.8%	1.0%	1.8%	10.5%	11.4%	1.0%	0.0%
classification accuracy	93.5%	94.4%	93.5%	88.0%	93.5%	89.5%	89.8%	88.9%	98.2%	97.2%

Table 3: Test results on radius fracture detection.

	left	right	overall
fracture detection rate	80.0%	84.6%	82.6%
false alarm rate	15.2%	19.5%	17.6%
classification accuracy	78.8%	75.6%	77.0%

CONCLUSION

This paper presented our research work on the detection of bone fractures in x-ray images. A suite of methods that combine different features and classification techniques have been developed and tested on detecting femur fractures. Although the fracture detection rate of individual classifier is not very high, they can complement each other in fracture detection. As a result, by combining the individual classifiers, both fracture detection rate and classification accuracy are improved significantly. Preliminary test results are also reported on the detection of radius fractures near the wrist, with encouraging results. We expect the combined approach to be applicable and useful for improving the fracture detection rate and classification accuracy of radius fractures as well.

In extracting the features for classification, adaptive sampling is used, which can adapt to the variation of size over different images. Another advantage of this adaptive sampling method is that it requires the extraction of only approximate bone contours. Therefore, it can also tolerate slight variation of shape over different patients, and does not require very accurate extraction of the bone contours.

Our continuing research objective is to study how strongly the presence of growth plates of the radius bone (Fig. 8) affects fracture detection and classification performance. Growth plates are a feature of the natural growing process of the radius, which should not be classified as fractures. Nevertheless, in some cases, growth plates can be presence together with fractures of the radius further away from the wrist (Fig. 9). These types of fractures are, however, easy to detect especially from the x-ray images of the side view of the wrist.

Our next target is to develop a prototype system for field test in the hospital. Although our laboratory tests indicate that the combined approach can detect all the fractures in the test set with only 10% false alarm rate, it will be interesting to see how well the system performs in field test.

Fig. 8: Healthy radius with growth plate.

Fig. 9: Fractured radius with growth plate.

ACKNOWLEDGMENTS

This research is supported by NMRC/0482/2000.

REFERENCES

[1] A. C. Bovik, M. Clark, and W. Geisler. Multichannel texture analysis using localized spatial filter. *IEEE Trans. PAMI*, 12(1):55–73, 1990.

[2] J. C. Buckland-Wright, J. A. Lynch, J. Rymer, and I. Fogelman. Fractal signature analysis of macroradiographs measures trabecular organization in lumbar vertebrae of postmenopausal women. *Calcif. Tissue Int.*, 54:106–112, 1994.

[3] P. Caligiuri, M. L. Giger, and M. Favus. Multifractal radiographic analysis of osteoporosis. *Medical Physics*, 21:503–508, 1994.

[4] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. European Conference on Computer Vision*, 1998.

[5] C. Cortes and V. N. Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 1995.

[6] G. R. Cross and A. K. Jain. Markov random field texture models. *IEEE Trans. PAMI*, 5:25–39, 1983.

[7] J. D. Furst and S. M. Pizer. Marching optimal-parameter ridges: An algorithm to extract shape loci in 3D images. In *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (LNCS 1496)*, pages 780–787, 1998.

[8] W. G. Geraets, P. V. der Stelt, and P. J. Elders. The radiographic trabecular bone pattern during menopause. *Bone*, 14:859–864, 1993.

[9] D. Guo and P. Richardson. Automatic vessel extraction from angiogram images. In *IEEE Conf. on Computers in Cardiology*, pages 441–444, 1998.

[10] IOF. The facts about osteoporosis and its impact. International Osteoporosis Foundation, www.osteofound.org/press_centre/fact_sheet.html, 2002.

[11] J. J. Kaufman, A. Chiabrera, M. Hatem, N. Z. Hakim, M. Figueiredo, P. Nasser, S. Lattuga, A. A. Pilla, and R. S. Siffert. A neural network approach for bone fracture healing assessment. *IEEE Engineering in Medicine and Biology*, 9(3):23–30, 1990.

[12] P. M. Khosrovi, A. J. Kahn, H. K. Genant, and S. Majumdar. In L. Raisz, editor, *American Society for Bone and Mineral Research*, page S156. Mary Ann Liebert, Inc., Kansas, USA, 1994.

[13] T. Link, S. Majumdar, W. Konermann, N. Meier, J. Lin, D. Newitt, X. Ouyang, P. Peters, , and H. Genant. Texture analysis of direct magnification radiographs of vertebral specimens: Correlation with bone mineral density and biomechanical properties. *Acad. Radiol.*, 4:167–176, 1997.

[14] J. A. Lynch, D. J. Hawkes, and J. C. Buckland-Wright. Analysis of texture in macroradiographs of osteoarthritic knees using the fractal signature. *Phys. Med. Biol.*, 36(6):709–722, 1991.

[15] S. Majumdar, J. Lin, T. Link, J. Millard, P. Augat, X. Ouyang, D. Newitt, R. Gould, M. Kothari, and H. Genant. Fractal analysis of radiographs: Assessment of trabecular bone structure and prediction of elastic modulus and strength. *Medical Physics*, 26(7):1330–1340, 1999.

[16] A. Matani, T. Okamoto, and K. Chihara. Evaluation of a trabecular structure using a multiresolution analysis. In *Proc. Annual Int. Conf. of IEEE Engineering in Medicine and Biology Society*, pages 632–633, 1998.

[17] A. Materka, P. Cichy, and J. Tuliszkiewicz. Texture analysis of x-ray images for detection of changes in bone mass and structure. In M. K. Pietikäinen, editor, *Texture Analysis in Machine Vision*. World Scientific, 2000.

[18] X. Ouyang, S. Majumdar, T. M. Link, P. A. Y. Lu, J. C. Lin, D. C. Newitt, and H. K. Genant. Morphometric texture analysis of spinal trabecular bone structure assessed using orthogonal radiographic projections. *Medical Physics*, 25:2037–2045, 1998.

[19] D. M. Ryder, S. L. King, C. J. Olliff, and E. Davies. A possible method of monitoring bone fracture and bone characteristics using a non-invasive acoustic technique. In *Proc. Int. Conf. on Acoustic Sensing and Imaging*, pages 159–163, 1993.

[20] A. Schwaighofer. SVM Toolbox for Matlab, www.cis.tugraz.at/igi/aschwaig/software.html.

[21] V. R. Singh and S. K. Chauhan. Early detection of fracture healing of a long bone for better mass health care. In *Proc. Annual Int. Conf. of IEEE Engineering in Medicine and Biology Society*, pages 2911–2912, 1998.

[22] T. P. Tian, Y. Chen, W. K. Leow, W. Hsu, T. S. Howe, and M. A. Png. Computing neck-shaft angle of femur for x-ray fracture detection. In *Proc. Int. Conf. on Computer Analysis of Images and Patterns (LNCS 2756)*, pages 82–89, 2003.

[23] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.

[24] V. N. Vapnik. *Statistical Learning Theory*. Wiley, Chichester, GB, 1998.

[25] J. Veenland, T. Link, W. Konermann, N. Meier, J. Grashuis, and E. Gelsema. Unraveling the role of structure and density in determining vertebral bone strength. *Calcif. Tissue Int.*, 61:474–479, 1997.

[26] D. W.-H. Yap, Y. Chen, W. K. Leow, T. S. Howe, and M. A. Png. Detecting femur fractures by texture analysis of trabeculae. In *Proc. Int. Conf. on Pattern Recognition*, 2004.