Detecting and Classifying Linear Structures in Mammograms Using Random Forests

Michael Berks, Zezhi Chen, Sue Astley, and Chris Taylor

Imaging Science and Biomedical Engineering, School of Cancer and Enabling Sciences, University of Manchester, Oxford Road, Manchester, M13 9PT, UK
michael.berks@manchester.ac.uk

Abstract. Detecting and classifying curvilinear structure is important in many image interpretation tasks. We focus on the challenging problem of detecting such structure in mammograms and deciding whether it is normal or abnormal. We adopt a discriminative learning approach based on a Dual-Tree Complex Wavelet representation and random forest classification. We present results of a quantitative comparison of our approach with three leading methods from the literature and with learning-based variants of those methods. We show that our new approach gives significantly better results than any of the other methods, achieving an area under the ROC curve $A_z = 0.923$ for curvilinear structure detection, and $A_z = 0.761$ for distinguishing between normal and abnormal structure (spicules). A detailed analysis suggests that some of the improvement is due to discriminative learning, and some due to the DT-CWT representation, which provides local phase information and good angular resolution.

Keywords: mammography**,** linear structure detection, spiculated lesions, classification, regression, random forests, dual-tree complex wavelet.

1 Introduction

The detection and classification of curvilinear structure is important in many image interpretation tasks – from roads and rivers in aerial images, through cracks in manufactured components, to blood vessels and ducts in medical images. Often the detection task is difficult because the structures of interest appear at relatively low contrast against a cluttered background. Classification is of interest when it is important to distinguish between subtly different structures which may be present within the same image – for example, rivers and roads. In this paper we focus on the challenging problem of detecting and classifying curvilinear structures in x-ray mammograms, but the methods we describe are equally applicable to problems in other domains.

We adopt a discriminative learning approach, applied to a rich description of local image structure, both to distinguish between curvilinear structure and background clutter, and to classify detected structures into different types. This approach achieves high detection accuracy by taking account of the domain-specific characteristics of both the structures of interest and the background clutter, whilst capturing subtle distinctions between different types of curvilinear structure. To describe local image

G. Székely and H.K. Hahn (Eds.): IPMI 2011, LNCS 6801, pp. 510–524, 2011.

structure we use the Dual-Tree Complex Wavelet Transform (DT-CWT) [1], whilst for detection and classification we use random forest (RF) classification [2].

In the remainder of the paper we provide more detailed motivation, explain the details of our approach and present the results of an experimental evaluation using both synthetic and real data. We show that, overall, our approach is significantly better than the current state-of-the-art, and that we can distinguish effectively between curvilinear structures associated with malignancy (spicules) and those associated with normal structure (vessels, ducts etc).

2 Background

2.1 Detecting Abnormalities in Mammograms

Breast screening programmes using x-ray mammography have been deployed widely to detect early signs of cancer. The use of computer-aided detection systems to support breast radiology experts is also widespread. In mammograms, the projection of a complex network of vessels, ducts and connective tissue in the breast results in images that are rich in linear structures at a variety of scales and orientations. Mammographic signs of cancer include: the presence of a radio-opaque mass, often with associated radiating curvilinear structures (spicules); the presence of microcalcifications (clusters of highly radio-opaque 'dots'); and the presence of abnormal or abnormally organised curvilinear structures (architectural distortion – AD). Fig 3(a) shows an approximately 4 x 4 cm region extracted from a mammogram, centred on a malignant tumour, and demonstrates a central mass and radiating spicules, along with normal curvilinear structures (vessels, ducts, glandular tissue etc). The signs of abnormality often appear in conjunction, but when only AD is present, detection of the abnormality is extremely difficult. It has been estimated that approximately a third of all cancers missed during mammographic screening present as AD [3], and it has been reported that computer-aided detection systems do not yet detect AD with adequate sensitivity or specificity [4].

Previous attempts at detecting both patterns of spicules associated with malignant masses and more general distortions in which no focal mass is visible have employed a two stage approach comprising i) the detection of all curvilinear structures in a specified region, ii) an analysis of the orientation pattern of the detected structures to determine whether abnormal patterns are present [5-7]. The work we report here seeks to provide a firmer foundation for this approach by improving the detection of curvilinear structures and identifying those structures likely to be (abnormal) spicules.

2.2 Detecting and Classifying Curvilinear Structures

There is a substantial literature on the detection of curvilinear structure in mammograms which we will review briefly. There is also a more general body of work on the detection of curvilinear structure which space does not allow us to cover, though most of the relevant approaches have been applied to mammograms.

Dixon and Taylor described a curvilinear structure detector based on comparing the integrated intensity along oriented lines (generally referred to as Linop) [8], which Parr et al. applied to mammograms [6]. Cerneaz and Brady [9] described an approach

based on estimating the second derivatives of the mammographic image surface. Karssemeijer and te Brake described an approach based on multiscale directional second-order Gaussian derivatives [5]. Zwiggelaar et al. compared all these methods using simulated mammographic images and reported that Linop gave the best results in terms of area under the receiver operating characteristic (ROC) curve [10]. More recently, researchers have recognised the importance of using local phase information, particularly in distinguishing between strong edges and genuine curvilinear structure. Rangayyan and Ayres described an approach based on Gabor filters [7]. Schenk and Brady [11] and McLoughlin et al. [12] used sets of steerable complex filters [13] at multiple scales to compute local energy, orientation and phase at each pixel. Wai et al. [14] used the monogenic signal as more efficient way of calculating local phase and orientation at multiple scales, detecting curvilinear structure using amplitude-weighted local phase congruency. Overall, the conclusion that can be drawn from the literature is that the detection of curvilinear structure benefits from access to local phase and magnitude information at multiple scales.

The literature on classifying curvilinear structures in mammograms is much more limited. We are aware of the work of Zwiggelaar et al. [10], which demonstrated the feasibility of distinguishing between different types of structure using cross-sectional profiles obtained from manually annotated curvilinear structures, but did not obtain very satisfactory results when the method was fully automated. We recently reported preliminary classification (but not detection) results using our current approach [15].

2.3 Dual-Tree Complex Wavelet Transform

Wavelet transforms have been used extensively in image processing and analysis to provide a rich description of local structure. The dual-tree complex wavelet transform has particular advantages because it provides a directionally selective representation with approximately shift-invariant coefficient magnitudes and local phase information [1]. The DT-CWT combines the outputs of two discrete transforms, using real wavelets differing in phase by 90°, to form the real and imaginary parts of complex coefficients. For 2-D images, it produces six directional sub-bands, oriented at ±15º, ±45º and ±75º, at each of a series of scales separated by factors of two. The DT-CWT is less expensive to compute and provides a richer description (magnitude and phase at six orientations rather than one) than the monogenic signal used by Wai et al. [14].

2.4 Random Forest Classification and Regression

Random forest classifiers have become popular because of their ease of use (rapid training, no critical parameters to select), resistance to overtraining, and near state-of-the-art performance. Given a set of training data consisting of N samples, each of which is a D-dimensional feature vector labelled as belonging to one of C classes, a random forest comprises a set of tree predictors constructed from the training data [2]. Each tree in the forest is built from a bootstrap sample of the training data (that is, a set of N samples chosen randomly, with replacement, from the original data). The trees are built using a standard classification and regression tree (CART) algorithm; however, rather than assessing all D dimensions for the optimal split at each tree node, only a random subset of $d < D$ dimensions are considered. The trees are built to

full size (i.e. until a leaf is reached containing samples from only one class) and do not need to be pruned (although they can be for the sake of efficiency).

During classification, an unseen feature vector is classified independently by each tree in the forest; each tree casts a unit class vote, and the most popular class can be assigned to the input vector. Alternatively, the proportion of votes assigned to each class can be used to provide a probabilistic labelling of the input vector. Random forests are particularly suited to learning non-linear relationships in high-dimensional multi-class training data, and have been shown to perform as well as classifiers such as Adaboost or support vector machines, whilst being more efficient to compute [2].

Similarly, by building regression rather than classification trees, a random forest can be constructed to learn the relationship between patterns of high-dimensional training data and a single output variable.

3 Detecting Curvilinear Structures

3.1 Detection as Learning

We have argued for the rich description of local structure provided by the DT-CWT. We now have to decide how to use this information to compute a single measure of curvilinear structure probability. One solution would be to select the maximum of the six oriented sub-band coefficients at each scale, and combine them across scale in a measure of phase congruency, as in the monogenic signal based method of Wai et al. [14], but this would discard potentially useful information. Instead, we pose the problem as a classification task in which we attempt to learn the patterns of DT-CWT coefficients associated with pixels belonging to either linear structures or background. We construct a feature vector characterising each pixel, by sampling DT-CWT coefficients (in phase/magnitude form) from the six oriented sub-bands in each of the s finest decomposition scales from a $w \times w$ neighbourhood centred on the pixel. This involves interpolating the coefficients obtained at coarse scales to provide a response at every pixel using the band-pass interpolation method of Anderson et al. [16]. To improve the rotational symmetry of the coefficients, we apply the transformations recommended by Kingsbury [17]. That is, an additional set of filters are used to reduce the wave frequencies of the 45° and 135° sub-bands so that they lie closer to the 15°, 75°, 105° and 165° bands. The six sub-bands are then multiplied by $\{i, -i, i, -1, 1, -1\}$ respectively, so that the phase at the centre of the impulse response of each wavelet is zero. Finally, to achieve 180° rotational symmetry, we replace any coefficient with negative imaginary part with its complex conjugate.

By forming such a feature vector at each pixel in a training set of ground-truth labelled images, we can obtain a large set of data which can be used to train a random forest classifier to differentiate between linear structures and background. In constructing the best detector, this discriminative approach takes account of the characteristics of the background pixels as well as the curvilinear structure pixels.

3.2 Orientation as Regression

In addition to detecting curvilinear structure in mammograms, for many subsequent analysis tasks (e.g. detecting abnormal patterns of structure indicative of disease) it is

equally important that we can make accurate estimates of orientation. As with detection, rather than doing this prescriptively, we pose the problem as a regression task in which we learn how patterns of DT-CWT coefficients correspond to structures at varying orientations. Feature vectors of DT-CWT coefficients are formed as in section 2.1, with the difference that the image ground truth specifies the orientation of a pixel as opposed to its class membership. A random forest regressor can then be trained to predict the orientation of any pixel given its DT-CWT representation.

3.3 Training on Synthetic Data

To apply our learning algorithm to mammograms we require a precise binary labelling of curvilinear structure and background pixels. Given the complexities of structure typically found in mammograms, such data are impractical to obtain. Our solution is to train a classifier on synthetic images containing local structure that closely resembles that which we expect to find in real mammograms. We form these synthetic images by pre-processing regions of real mammograms to remove existing linear structure, whilst maintaining both the coarse scale texture properties of the region and the local image noise across all frequencies, before superimposing synthetic curvilinear structures on the regions. We have used straight lines, each defined by its orientation, width, peak intensity and cross-sectional intensity profile. For each property, we define a probability density function from which samples can be drawn, allowing arbitrarily large numbers of lines to be generated. An example of a synthetic image is shown in Fig 1(a). More details are given in section 5.1.

Given a set of synthetic images generated as described above, we can train a random forest classifier. The literature suggests (and this is borne out by our experience) that the best performance is obtained by training with a balanced dataset – in this case with the same number of foreground (curvilinear structure) and background examples. To achieve this we randomly sample feature vectors from pixels lying on the centrelines of synthetic curvilinear structures, and randomly sample the same number of feature vectors from background pixels. Having trained a classifier using the synthetic data, we can classify feature vectors extracted about every pixel in a synthetic or real mammogram image to obtain a line probability image (using the probabilistic labelling scheme as opposed to a hard binary classification). Example results are shown in Fig 1(b)-(h) and Fig 3 (b)-(h).

To construct the orientation regression random forests, we sample points randomly from line pixels in the synthetic images (background pixels are not used because the orientation is undefined).

4 Classifying Curvilinear Structure

The learning approach described above can also be used to differentiate between different kinds of curvilinear structure. The hypothesis is that the cross-sectional intensity profiles of structures differ in systematic ways between types of mammographic structure (as suggested by Zwiggelaar et al. [10]), and that profile shape is effectively captured by the DT-CWT coefficients – particularly in the phase components. In the experiments described below we concentrated on the task of distinguish-

ing between spicules, which are a sign of malignancy, and other curvilinear structures. Although it was not realistic to create a training set of mammogram images with all spicules annotated, we were able to obtain expert annotations of a reasonably large number of spicules in a set of mammogram patches containing malignant tumours (see section 5.1 for details). These annotations were used to select pixels for the positive training set. To create a balanced training set we sampled feature vectors from the same number of pixels in a set of normal mammogram patches, such that the distribution of curvilinear structure probabilities was the same as for the spicule training set. Using this balanced training set, we built a random forest classifier to perform spicule/non-spicule classification.

5 Experimental Evaluation

We have conducted a systematic evaluation of the performance of our method for curvilinear structure detection and classification, using both synthetic and real mammogram data, comparing seven competing approaches:

- **DT-CWT/RF:** the method described in this paper.
- **Monogenic:** the monogenic-signal phase congruency approach of Wai et al. [14].
- **Linop:** the Line operator method of Dixon and Taylor [8] and Parr et al. [6].
- **Gaussian:** the directional Gaussian 2^{nd} derivatives method employed by Karssemeijer and Te Brake [5].
- **Monogenic/RF:** the raw responses used in *Monogenic*, combined using RF classification.
- **Linop/RF:** the raw responses used in *Linop*, combined using RF classification.
- **Gaussian/RF:** the raw responses used in *Gaussian*, combined using RF classification.

Monogenic, Linop and Gaussian are representative of the current state of the art in line detection. Monogenic/RF, Linop/RF and Gaussian/RF are natural variants, in which the intermediate multiscale responses previously used to construct the detection outputs are instead combined to given feature representation at each pixel that can subsequently classified using a random forest. These learning variants were developed for two reasons: firstly, when analyzing quantitative results for detection performance, they allow us to decouple the effect of random forest learning from the effect due to the type of representation used. Secondly, unlike their original variants, Monogenic/RF, Linop/RF and Gaussian/RF can be used in the spicule classification experiment described in 5.3.

In what follows, we present both qualitative and quantitative results for detection and classification for each method outlined above.

5.1 Data

Synthetic Data. We used two sets of synthetic images containing curvilinear structures, with known ground truth, to train and test curvilinear structure detection. We

randomly extracted 4 × 4 cm (512 × 512 pixel) mammographic backgrounds with 256 grey-levels 72 (30) normal mammograms for the training (test) set, resulting in 10460 training patches and 4903 test patches, from which naturally occurring linear structures were filtered out. Lines were added to these backgrounds, with parameters drawn from the following distributions: orientation [0, π] uniform; width [4, 16] pixels uniform; peak intensity [1,255] grey-levels (relative to images scaled 0 – 255) from an exponential distribution with half-width 4 grey-levels; profile shape β determined by the equation $\beta = \alpha + (1- \alpha) \sin x$ for offsets $x \in (0,\pi)$, where the 'squareness' parameter α determines how close the shape is to a pure half-cycle sin or a rectangle and is drawn uniformly from [0,1]. The range of widths, distribution of contrasts (with low contrast lines much more likely than high contrast lines) and variability in profile shape were chosen to mimic what is found in real mammograms.

During training, backgrounds were randomly sampled from the 10460 training patches and a single line was added to the centre of the patch. These images were produced 'on-the-fly' during each tree-building step of random forest construction as described in section 5.2 and no permanent set was maintained.

For testing, 100 backgrounds were randomly selected from the test patches. To each, multiple lines were added sequentially, with the number and position of lines varying randomly. An example of a synthetic image is shown in Fig 1(a).

Note that all synthetic lines used were straight lines. We conducted experiments explicitly using curved structures, however as there was no performance difference between training on curved or straight lines when detecting curved lines, it was decided that including curves was unnecessary.

Real Mammogram Data. We used real mammogram patches to illustrate qualitative results for curvilinear structure detection and to train and test spicule/non-spicule classification. Data were taken from a sequential set of 84 abnormal mammograms with biopsy-proven malignancy, drawn from a screening population (Nightingale Breast Centre, South Manchester University Hospitals Trust, UK), and from a set of 89 normal mammograms of the contralateral breasts of the same individuals (where disease was radiologically confirmed to be confined to one breast). All mammograms were digitised to a resolution of 90μm, using a Vidar CADPRO scanner. A 4×4 cm patch was extracted around each abnormality, and a similar patch was sampled randomly from each of the normal mammograms; examples are shown in Fig 4. For each abnormal patch an expert radiologist manually annotated some (though not necessarily all) of the spicules associated with the abnormality, resulting in a total of 555 spicule annotations. The expert spicule annotations for the abnormal images were used as a basis for selecting spicule pixels, though they were not sufficiently accurate to be used directly. To refine the annotations, we initialised a snake [18] using each original annotation, and iterated it to convergence, using evidence from the linear structure probability image. We note that a similar technique has recently been published in detail by Muralidhar et al. [19]. As a result, the 555 refined spicule annotations identified a set of 36,514 spicule pixels. As outlined in section 4, we sampled the same number of pixels from the normal images, such that the detection probability distributions for the spicule and non-spicule samples were the same.

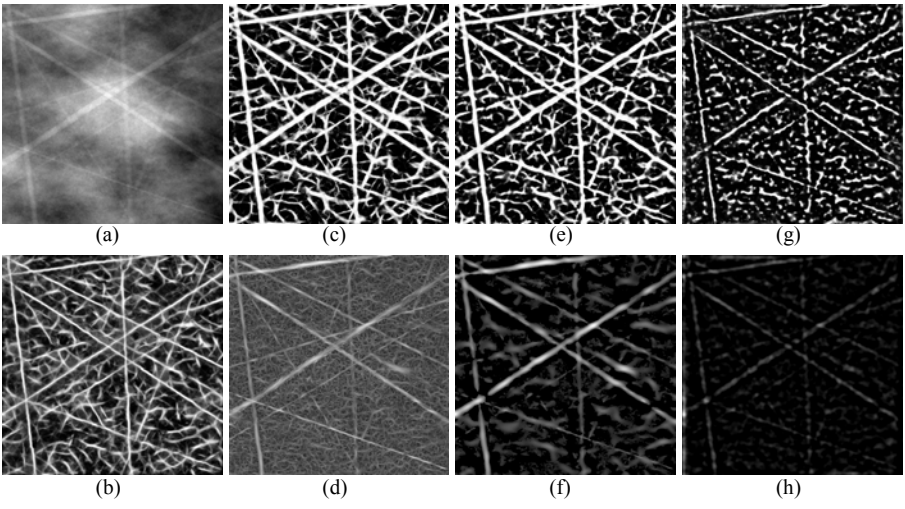

Fig. 1. a) Synthetic test image; **b) - h)** results of line detection algorithms (see text for details): **b)** *DT-CWT/RF*; **c)** *Linop/RF*; **d)** *Linop*; **e)** *Gaussian/RF*; **f)** *Gaussian*; **g)** *Monogenic/RF*; **h)** *Monogenic*

5.2 Curvilinear Structure Detection and Orientation Estimation

Curvilinear structure detection and orientation estimation was tested for all seven methods outlined at the beginning of section 5.

For the four learning methods (DT-CWT/RF, Monogenic/RF, Linop/RF, Gaussian/RF), we first constructed random forests to classify between structure and background pixels and to compute an estimate of orientation at each pixel.

All forests were constructed with 200 trees and $d = \sqrt{D}$, following published guidelines [2]. However, rather than using a single set of training data from which samples were drawn with replacement (i.e. bootstrapping), we used our method for randomly generating unique synthetic lines images (as described in section 5.1) to construct a new training sample at each tree-building step. For detection, each sample comprised 100k curvilinear structure pixels and 100k background pixels, whilst for orientation regression we used 200k pixels all belonging to curvilinear structure.

Thus for any given representation (*DT-CWT, Monogenic, Linop, Gaussian*) and forest (classification, regression) we applied the following scheme:

1. Generate a new synthetic line image with known ground truth
2. Extract feature vectors for a random set of pixels in the image
3. Repeat 1 & 2 until training sample complete
4. Construct tree using the CART algorithm
5. Repeat 1 to 4 until 200 trees constructed

Details on the composition of feature vectors for each representation type are given below. Note for all methods, the number of scales used $\boldsymbol{s}$ and the neighbourhood size $\boldsymbol{w}$ were empirically tested to select the best combination for each method. In each

case, we tried s = 3, 4, 5, 6 and w = 1, 3. For reasons of space, results are shown only for the best combination in each method.

- **DT-CWT/RF:** images were decomposed using the DT-CWT to s scales. Neighbourhoods of interpolated phase and magnitude coefficients were extracted in each of the 6 oriented subbands producing a feature vector of $12\mathbf{w}^2\boldsymbol{s}$ elements.
- **Monogenic/RF:** the monogenic signal was extracted across s scales, with the initial wavelength set at λ= 4 pixels. Neighbourhoods of phase, amplitude and orientation values were computed giving a total feature size of 3 $\mathbf{w}^2\boldsymbol{s}$.
- **Linop/RF:** 8 oriented line filters were computed at each scale. Collecting neighbourhoods of the oriented filter responses produced $8\mathbf{w}^2\boldsymbol{s}$ elements in each feature vectors.
- **Gaussian/RF:** for the Gaussian 2nd derivative method, the three directional derivatives were applied to an image at each scale. The standard deviation of the smallest kernel was 1 pixel, subsequent kernels increased by a factor of 2 at each scale. As with *Monogenic/RF* this resulted in feature vectors with $3\mathbf{w}^2\boldsymbol{s}$ elements.

For testing, feature vectors for each representation were extracted at all pixels in the 100 synthetic test images. The classification and regression forests were then used to compute a line detection score (the proportion of votes for the curvilinear structure class) and orientation (the mean output of each regression tree, with appropriate angular wrapping at 0º and 180º) at each pixel. Example results of line detection are shown in Fig1 (b)-(e).

In addition to the four learning methods, the prescriptive variants of the Monogenic, Linop and Gaussian approaches were applied to the test images, example results of which are depicted in Fig 1 (f)-(h).

ROC curves for the seven methods tested are shown in Fig 2, using the known ground-truth for the test images to define pixels on the centrelines of curvilinear structures as foreground, and pixels lying outside the structures as background. Areas under the ROC curves and detection sensitivities at a fixed specificity of 90% are shown in table 1. For orientation, only pixels belonging to curvilinear structures (although not necessarily at the centerline) were included in the results. The absolute differences between the predicted and known orientations (with appropriate angle wrapping) were taken, and used to generate cumulative distribution functions of orientation errors for each method, as shown in Fig 2. The mean absolute errors of the estimated orientations are also included in Table 1.

These results show that the four learning methods perform significantly better than the three prescriptive methods (with the exception of orientation computation in *Monogenic/RF*). *DT-CWT/RF* is significantly the strongest performing for both line detection and orientation estimation, followed by *Linop/RF* then *Gaussian/RF*.

As expected, because *Linop*, of the three prescriptive methods, discards the highest proportion of filter responses, *Linop/RF* gains the most from training. This highlights the ability of the random forests to extract useful information from a rich local description of image content, and whilst we do not have a prescriptive variant to compare it to, we believe this shows the importance of training in maximizing the benefit of using the DT-CWT. We also note that the added information that can be gained from the DT-CWT representation results from an increase in the richness of

the local description of texture and is not due to increasing the support of the descriptor. Indeed, as described above we tested all representations over an equivalent range of filter scales so that the same image support was available to each method.

The orientation results for both *Monogenic* and *Monogenic/RF* were surprisingly poor. Counter-intuitively, visual analysis of the test outputs showed that the *Monogenic* methods performed particularly badly at the exact centerline of curvilinear structures, where an essentially random orientation appeared to be returned. This is in contrast to the other tested methods that, as expected, performed strongest along the centerlines of structures. Computing estimation errors at pixels belonging to structures but not lying on the centerline produces a small improvement in the results (mean absolute errors of 32.55º and 29.39º for the *RF* and prescriptive variant respectively), though even then performance lags behind the other tested methods and of course in a real image we do not know the precise location of structure centerlines. Determining why orientations are estimated so poorly by the monogenic signal at the centre of structures, where in theory the signal is strongest, may warrant further investigation.

Table 1. Line detection and orientation computation results

Line detection algorithm	ROC A_z	Sensitivity for fixed 90% specificity	Mean Absolute Error of orientation
DT-CWT/RF, $w = 1$, $s = 5$	0.923±0.00036	0.792	15.88
Linop/RF, $w = 1$, $s = 5$, 8 orientations	0.911±0.00048	0.757	19.35
Gaussian/RF, $w = 1$, $s = 4$, $\sigma_{min} = 1$	0.901±0.00048	0.731	21.37
Monogenic/RF, $w = 1$, $s = 4$, $\lambda = 4$	0.868±0.00045	0.643	33.73
Monogenic, $s = 3$, $\lambda = 4$	0.818±0.00055	0.547	30.86
Gaussian, $s = 4$, $\sigma_{min} = 1$	0.813±0.00053	0.533	24.27
Linop, $s = 5$, 8 orientations	0.737±0.00060	0.413	29.32

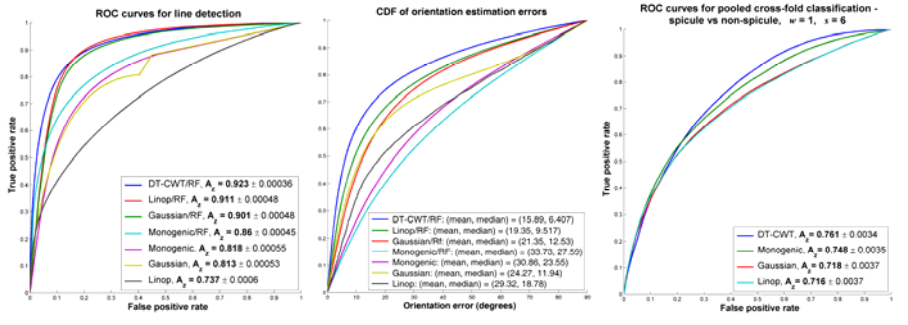

Fig. 2. *left*: Receiver operating characteristic (ROC) curves for different curvilinear structure detection methods; *centre*: Cumulative distribution functions (CDF) of errors in orientation estimation for the different methods; *right*: ROC curves for spicule classification.

To show qualitative results for real mammograms, the seven methods were applied to detect curvilinear structures and estimate orientation for the malignant regions described in section 5.1. Example detection results are shown in Fig 3. Assessing the results visually, it would appear that the outputs of the learning based methods (and particularly *DT-CWT/RF*, *Linop/RF* and *Gaussian/RF*) are similar to the output of synthetic data whilst capturing the key curvilinear structures in the real regions. This would suggest our model for producing synthetic lines generates good data from which to train random forests for real images. Of course validating this hypothesis is important and we are currently working on a quantitative evaluation of the methods in real data when used, for example, as part of a larger lesion detection scheme.

In terms of applying the learning methods to real images, it is worth noting how the methods scale with increasing image size – particularly above the point at which the set of feature vectors for all image pixels can be stored in working memory. For the DT-CWT, provided the initial decomposition can be stored in memory (which due to its low-redundant decimating construction is possible even for full size mammograms of the order 3000x2400 pixels) then interpolated coefficients can be efficiently sampled to generate feature vectors for block-wise regions of the image. Each block of feature vectors can be classified by the forest and cleared from working from memory storing only the output of the forest. In this way only a small overhead is introduced for block-wise classifying the image. However, for the other methods it becomes necessary to interleave the decomposition of the image with the sampling of feature vectors. For example, it may be necessary to apply the filters at a single scale, extract features for that scale for a particular block of the image, filter at the next scale and extract those features, and so on. Of course, when the complete feature vectors for a single block have been classified, the process repeats. Thus a large image may in

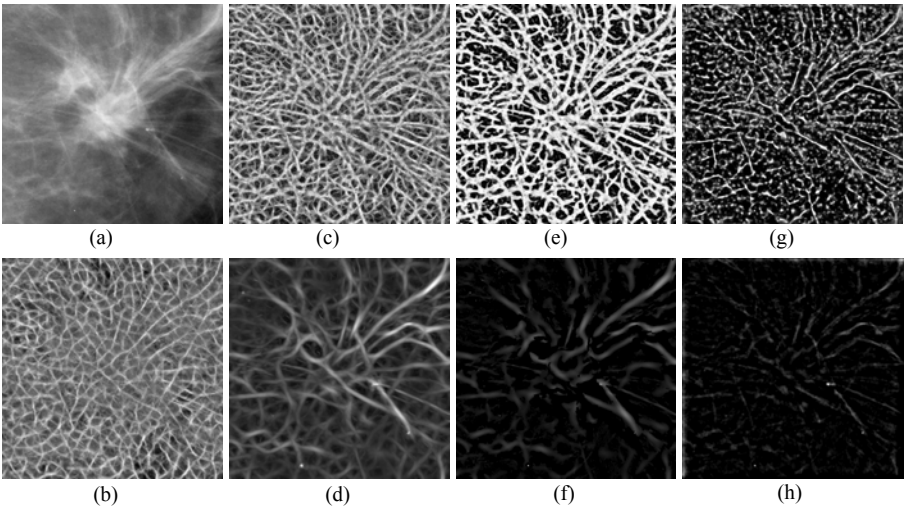

Fig. 3. a) Mammogram region containing malignant spiculated mass; **b) - h)** results of line detection algorithms (see text for details): **b)** *DT-CWT/RF*; **c)** *Linop/RF*; **d)** *Linop*; **e)** *Gaussian/RF*; **f)** *Gaussian*; **g)** *Monogenic/RF*; **h)** *Monogenic*

fact end up by decomposed many times over introducing a large computational overhead for block-wise processing. The point at which this cost occurs will depend on the size of the image and the type of representation used. Obviously the cost is worst for Linop which requires storing 8 (i.e. the number orientations) full copies of the image at each scale, compared to just 3 for the Gaussian and Monogenic methods.

5.3 Spicule Classification

The four learning-based methods were also applied to the problem of spicule/non-spicule classification. Feature vectors were formed as above, and random forest classifiers were trained using balanced spicule/non-spicule training data, as outlined in section 4. To make effective use of the available data, we used a 10-fold cross-validation design. The set of normal and abnormal regions were divided into 10 groups so that the total number of normal and spicule pixels in each group were as close as possible to a 10^{th} of the total and no two views from the same case were included in different groups. The samples in each group were then classified using a random forest trained on the samples from the remaining 9 groups. The classification results from each group were pooled to generate an unbiased class probability for each sampled pixel. These probabilities were used to compute an ROC curve for each training regime, and the area under the curve (A_z) was computed and used as a measure of classification performance. The ROC curves and A_z values for the three methods are shown in Fig 2 and Table 2. These results demonstrate a clear advantage for *DT-CWT/RF*. As might be expected, because the *Linop* and *Gaussian* representations do not really capture profile shape, they perform significantly worse that the two representations that include phase.

Table 2. Results for spicule classification

Feature representation for spicule classification	ROC A_z
DT-CWT/RF, $w = 1$, $s = 6$, all orientations	0.761±0.0034
Monogenic/RF, $w = 1$, $s = 5$, 8 orientations	0.748±0.0035
Gaussian/RF, $w = 1$, $s = 5$, $\sigma_{min} = 1$	0.718±0.0037
Linop/RF, $w = 1$, $s = 5$, $\lambda = 4$	0.716±0.0037

In addition to computing a class vote for spicule membership at only those pixels in the selected training sets, the forests we have constructed can be used to compute results for whole region in each cross-fold group. Typical qualitative *DT-CWT/*RF results for a normal and abnormal region are shown in Fig 4. In the left column, the original regions are shown. The spiculations of the mass are clear and well defined, particularly to the south-east of the central mass. In the normal region, there are a set of structures that intersect in an approximate radial pattern that may trigger a feature detector erroneously. In the right column, the predicted spicule class membership is shown as hue varying from cyan (normal) to pink (spicule), modulated by the output of the *DT-CWT/RF* detection method. Note how the linear structures in the region of the mass are deemed highly likely to be spicules, whilst those in the normal region are not. This shows excellent promise as a means of providing a relevance measure to methods for abnormality detection.

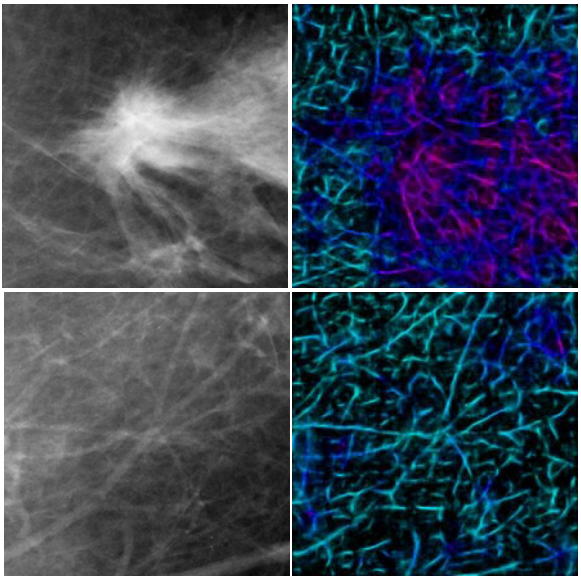

Fig. 4. Left column: an malignant (top) and normal (bottom) region; right column: spicule probability depicted as hue from cyan (normal) to pink (spicule), modulated by strength line detection output from the *DT-CWT* method

6 Discussion and Conclusions

We have presented a discriminative learning-based approach to the detection and classification of curvilinear structures, based on a combination of DT-CWT representation of local structure and random forest classification. We have applied the method to the challenging problem of detecting and estimating the orientation of curvilinear structures in mammograms and distinguishing between normal and abnormal structures. The results of our experimental evaluation are extremely encouraging, and represent a significant improvement over the current state of the art.

We have also introduced learning-based variants of three existing methods, demonstrating that whilst learning accounts for a significant part of this improvement, the choice of representation is also important and will have a different effect on performance depending on the task in hand. For example, constructing a representation based on the raw responses to Linop filters produces features that are excellent for estimating structure orientation but provide less information for determining structure shape and thus type. Conversely, features formed from the monogenic signal are good at determining structure type – most likely because of the inclusion of the phase measure – whilst they perform relatively poorly at detection and orientation estimation. For these reasons, it seems fair to conclude that the DT-CWT provides the best all round representation. It produced the strongest performance for all three tasks (curvilinear structure detection, orientation estimation and spicule classification). Moreover, as discussed in section 5.2, of all the methods, the DT-CWT incurs the least overhead

when working with full-size real images that require block-wise classification/regression. For example, initial tests show that the structure detection and orientation regression can be performed on a full-size (~3000 x 2400 pixels) mammogram in ~1hr 30mins.

Our next goal is to show that improving the individual steps of curvilinear structure and orientation estimation result in a subsequent improvement for a high level task such as detecting patterns of spiculations indicative of disease. Moreover we hope to show that classification into structure type can further aid such tasks by focusing only (or at least more) on those structures most likely to be associated with disease.

Acknowledgements

We thank Nick Kingsbury for the DT-CWT Matlab toolbox. Mammograms were provided by the Nightingale Breast Centre, South Manchester University Hospitals Trust, UK and were annotated by Dr Caroline Boggis and Dr Rumana Rahim. This work was funded by EPSRC grant EP/E031307/1.

References

1. Kingsbury, N.: Complex wavelets for shift invariant analysis and filtering of signals. Applied and Computational Harmonic Analysis 10(3), 234–253 (2001)
2. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
3. Burrell, H.C., Evans, A.J., Robin, A., Wilson, M., et al.: False-negative breast screening assessment: what lessons can we learn? Clin. Radiol. 56(5), 385–388 (2001)
4. Baker, J.A., Rosen, E.L., Lo, J.Y., Gimenez, E.I., et al.: Computer-Aided Detection (CAD) in Screening Mammography: Sensitivity of Commercial CAD Systems for Detecting Architectural Distortion. Am. J. Roentgenol. 181(4), 1083–1088 (2003)
5. Karssemeijer, N., te Brake, G.M.: Detection of stellate distortions in mammograms. IEEE Trans. Med. Im. 15(5), 611–619 (1996)
6. Parr, T.C., Taylor, C.J., Astley, S.M., Boggis, C.R.M.: Statistical modeling of oriented line patterns in mammograms. In: Proc. SPIE Medical Imaging, vol. 3034, pp. 44–55 (1997)
7. Rangayyan, R., Ayres, F.: Gabor filters and phase portraits for the detection of architectural distortion in mammograms. Medical and Biological Engineering and Computing 44(10), 883–894 (2006)
8. Dixon, R.N., Taylor, C.J.: Automated asbestos fibre counting. In: Institute of Physics Conference Series, vol. 44, pp. 178–185 (1979)
9. Cerneaz, N., Brady, M.: Finding curvilinear structures in mammograms. In: Computer Vision, Virtual Reality and Robotics in Medicine, pp. 372–382 (1995)
10. Zwiggelaar, R., Astley, S.M., Boggis, C.R.M., Taylor, C.J.: Linear structures in mammographic images: detection and classification. IEEE Trans. Med. Im. 23(9), 1077–1086 (2004)
11. Schenk, V., Brady, M.: Finding CLS Using Multiresolution Orientated Local Energy Feature Detection. In: Proc. 6th International Workshop on Digital Mammography, Bremen, Germany (2002)
12. McLoughlin, K.J., Bones, P.J., Kovesi, P.: Connective tissue representation for detection of microcalcifications in digital mammograms. In: Proc. SPIE Medical Imaging, vol. 4684, pp. 1246–1256 (2002)

13. Freeman, W.T., Adelson, E.H.: The Design and Use of Steerable Filters. IEEE Transactions on Pattern Analysis and Machine Intelligence 13, 891–906 (1991)
14. Wai, L.C.C., Mellor, M., Brady, J.M.: A Multi-resolution CLS Detection Algorithm for Mammographic Image Analysis. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 865–872. Springer, Heidelberg (2004)
15. Chen, Z., Berks, M., Astley, S.M., Taylor, C.J.: Classification of Linear Structures in Mammograms Using Random Forests. In: Martí, J., Oliver, A., Freixenet, J., Martí, R. (eds.) IWDM 2010. LNCS, vol. 6136, pp. 153–160. Springer, Heidelberg (2010)
16. Anderson, R., Kingsbury, N., Fauqueur, J.: Coarse-level object recognition using interlevel products of complex wavelets. In: Proc. IEEE International Conference on Image Processing, pp. 745–748 (2005)
17. Kingsbury, N.: Rotation-Invariant Local Feature Matching with Complex Wavelets. In: Proc. European Conference on Signal Processing, EUSIPCO (2006)
18. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision 1(4), 321–331 (1988)
19. Muralidhar, G.S., Bovik, A.C., Giese, J.D., Sampat, M.P., et al.: Snakules: A Model-Based Active Contour Algorithm for the Annotation of Spicules on Mammography. IEEE Trans. Med. Im. 29(10), 1768–1780 (2010)