Experiments in lung cancer nodule detection using texture analysis and neural network classifiers *

G. S. Cox, F. J. Hoare, G. de Jager (Member IEEE)
Department of Electrical Engineering
University of Cape Town
South Africa

Abstract

This paper discusses experiments in automatic lung cancer nodule detection in digitised chest x-rays. Texture features are extracted from subregions in an image. These features are used to train a neural network to classify the subregions as nodule tissue or otherwise. Two types of texture analysis, Haralick's co-occurrence matrices and Laws' texture measures, are used to generate texture feature vectors. Different neural network topologies, including back-propagation and learning vector quantisation networks, are used to evaluate the validity of the feature sets, and are investigated as the classification stage for nodule detection.

1 Introduction

The successful treatment of lung cancer is dependent on early diagnosis, which involves the identification of nodules in chest radiographs [1]. However, in many cases, observer error causes nodules to go undetected. If the radiograph is digitised, image processing techniques can be used to enhance the image, and pattern recognition could bring sites of nodules to the attention of the observer.

Many techniques have been proposed for the automatic detection of lung cancer nodules in chest x-rays [1–6]. Texture analysis could be considered as a means of segmenting the image into areas indicating nodules and other areas. Another approach would be to use texture features to classify possible sites of nodules that have been identified with previous processing.

This paper discusses experiments in the application of statistical texture features proposed by Haralick [7] and Laws [8]. Sets of features are extracted from chest x-rays of the different stages of primary lung cancer. Neural networks are used to evaluate feature sets and perform classification. *Learning vector quantisation* networks, and different variations of the *back-propagation* topology are used.

*Originally presented at the Third South African Workshop on Pattern Recognition. Pretoria, 1992

2 Statistical texture analysis

Current techniques describe texture using *statistical* or *structural* properties in the image. Structural techniques are less likely to be useful here because they rely on the derivation of a placement rule for texture primitives. Since anatomical objects fall somewhere between deterministic and stochastic texture, statistical texture analysis seems appropriate.

2.1 Haralick's texture features

Haralick *et al* [7] proposed *co-occurrence matrices* (or *grey tone spatial dependence matrices*) for texture analysis, based on the assumption that the texture-context information in an image is contained in the overall or average spatial relationship between the grey tones, or intensities.

2.1.1 Co-occurrence matrices

The co-occurrence matrix for distance d and angle θ is defined as follows:

> The matrix of relative frequencies CM_{ij}, with which two pixels separated by distance d and angle θ occur on an image where one has the grey level i, and the other has grey level j.

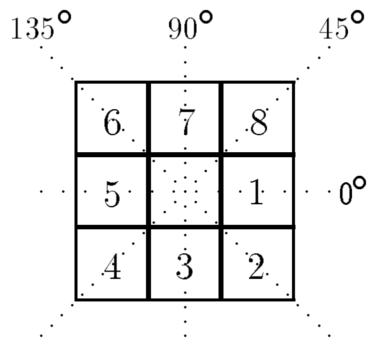


Figure 1: Definition of Haralick's angles

In other words the rows and columns of the co-occurrence matrix (CM) are indexed by the grey level range in an image. An image with n bits per pixel will have $G \times G$ co-occurrence matrices, where

$$G = 2^n \tag{1}$$

and $CM_{g_1 g_2}$ indicates the number of times grey levels g_1 and g_2 are found d pixels apart at an angle of θ. Haralick considers only the angles 0°, 45°, 90° and 135°. Figure 1 shows how these angles are defined. For example, consider the 4×4 image in figure 2(a)[1], where $g \in \{0, 1, 2, 3\}$. Two of the co-occurrence matrices for $d = 1$ are given in figure 2(b) and 2(c).

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

(a)

$$\begin{bmatrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{bmatrix} \qquad \begin{bmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

(b) $P_{135°}$ (c) $P_{0°}$

Figure 2: (a) Example image. Co-occurrence matrices for (b) $\theta = 135°$ and (c) $\theta = 0°$.

2.1.2 Statistical features

The texture features are obtained from statistics of the co-occurrence matrices. Haralick has defined 24 features including the *Angular Second Moment*, *Contrast*, *Correlation*, *Inverse difference moment* and *Entropy*. These are defined as follows:

Angular Second moment

$$f_1 = \sum_i \sum_j \{CM_{ij}\}^2 \tag{2}$$

Contrast

$$f_2 = \sum_{n=0}^{N_g - 1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{\substack{j=1 \\ |i-j|=n}}^{N_g} CM_{ij} \right\} \tag{3}$$

Correlation

$$f_3 = \frac{\sum_i \sum_j (ij) CM_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{4}$$

Inverse Difference Moment

$$f_5 = \sum_i \sum_j \frac{1}{1 + (1 - j)^2} CM_{ij} \tag{5}$$

Entropy

$$f_9 = -\sum_i \sum_j CM_{ij} \log(CM_{ij}) \tag{6}$$

These features are calculated for a region in the image and then constitute a feature vector that characterises the texture of that region.

2.2 Laws' texture features

Like Haralick's co-occurrence matrix features, Laws' features are a method of measuring texture energy in the spatial domain. Hsiao *et al* review Laws' features [9].

2.2.1 The microtexture feature plane

Laws defined 25 *microtexture masks* which are basically matched filters for certain common types of quasi-periodic texture in images. Figure 3 shows one of Laws' masks.

$$E5L5 = \begin{bmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Figure 3: One of Laws' microtexture feature masks

One of these masks is convolved with an image to create a *microtexture feature plane*. That is, for image f and mask h the microtexture feature plane x is given by

$$x(i, j) = h(i, j) * f(i, j) \tag{7}$$

Since the texture was assumed to be quasi-periodic, areas with high variance in x indicate positions where the microtexture mask, h, matched the local texture in f. The variance information is extracted from x in the form of *macrostatistical features*.

2.2.2 Macrostatistical features

Macrostatistical features are statistics obtained from windows in the microtexture feature plane. These windows are typically 15×15 or 31×31. The texture feature for point (i, j) on a microtexture feature plane

[1] Example taken from [7]

is calculated as the variance in a window centered on this point. The sample mean deviation, or ABSAVE, is an equivalent statistic that is more computationally efficient.

For a $2n+1$ by $2n+1$ window centered on (i,j), the ABSAVE is defined by:

$$s(i,j) = \frac{1}{(2n+1)^2} \sum_{k=i-n}^{i+n} \sum_{l=j-n}^{j+n} |x(k,l) - m(i,j)| \quad (8)$$

where the mean $m(i,j)$ is given by

$$m(i,j) = \frac{1}{(2n+1)^2} \sum_{k=i-n}^{i+n} \sum_{l=j-n}^{j+n} x(k,l) \qquad (9)$$

The ABSAVEs of each microtexture feature plane are Laws' *texture energy measures.*

3 Image acquisition and preprocessing

X-rays of two cases of lung cancer, in advanced and intermediate stages of nodule development, are used for these experiments. In each case the right and left lungs were considered separately.

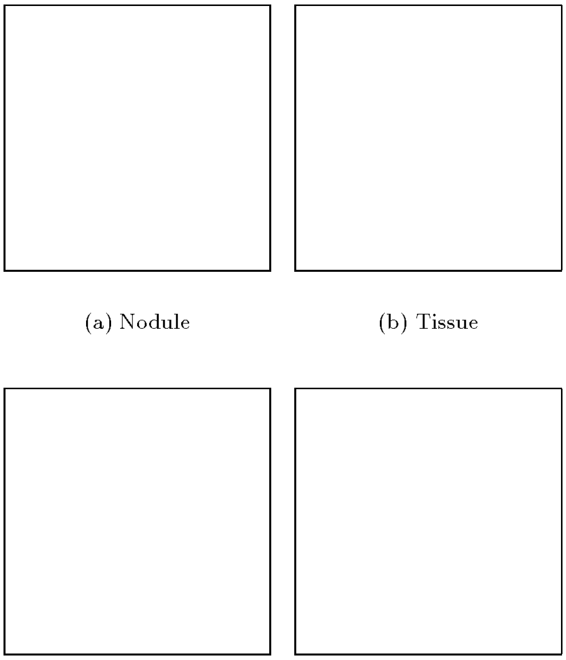
3.1 Digitising chest x-rays

The chest x-rays were digitised at 300 dpi and 8 bits per pixel. An OmniMedia XRS scanner was used at its lowest scan speed. This resulted in images of around 2000×3000 pixels. These images were binned with 2×2 and 4×4 superpixels, which created 10 bit and 12 bit images respectively. We found the 10 bit images the most useful, that is, the best trade-off between spatial and radiometric resolution.

3.2 Texture enhancement

In addition to the unprocessed images, texture analysis of preprocessed images is also considered. Texture information is present in the intensity gradient of small regions in the image. For this reason a Sobel edge detector, which gives an indication of gradient magnitude in the image, is used to enhance the texture information. The texture analysis techniques described before are then also applied to the Sobel preprocessed images. Subimages showing a nodule and background tissue, and subimages of their Sobel processed counterparts, are given in figure 4.

Therefore the entire test set of images consists of four images of right and left lungs, and Sobel edge detected versions of these images.



(a) Nodule (b) Tissue

(c) Sobel (d) Sobel

Figure 4: Images of a nodule and background tissue.

4 Feature extraction

Texture features are extracted from the test images. Sites of nodules, ribs and background tissue were manually classified. Haralick's and Laws' texture features are extracted for these sites and used as training and test sets for neural network classifiers.

4.1 Haralick's co-occurrence matrices

Co-occurrence matrices are not effective with low spatial, and high intensity resolution. Many intensity levels, and a small subimage, lead to a very sparse matrix. This is a problem with the images of original x-rays. We solve this problem by removing any additive offset from the subimage under consideration, and normalising the number of intensity levels to a constant. This approach also has the advantage of introducing additive invariance. We experiment with a 31×31 subimage, normalised to 64 intensity levels.

Co-occurrence matrices were generated for $d = 1$. Of the 24 features defined by Haralick, we use the five that are given in section 2.1.2.

4.2 Laws' texture features

All of the features generated using Laws' 25 microtexture masks are used. The shareware signal processing package, Khoros [10], is used to generate the micro-texture feature planes.

5 Classification

In order to classify the results of the texture analysis processes, various configurations of the back-propagation (BP) and learning-vector-quantisation (LVQ) neural network paradigms are investigated.

5.1 Back-propagation

The BP neural network is a feed forward neural network that is trained to find a mapping between the inputs to the network and the output of the network. This is accomplished by presenting preclassified input-output sets of data to the network iteratively and updating the connecting weights according to equation 10 [11,12].

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_j x_i \tag{10}$$

where η is a gain term, δ_j is an error term for node j, and x_i is the output from node i.

5.1.1 Variations of the standard BP algorithm

Several different variations of the basic equation 10 are implemented.

1. **Momentum:** A momentum term defined by equation 11 is added to equation 10.

$$\alpha(w_{ij}(t) - w_{ij}(t-1)) \tag{11}$$

   This has the effect of enforcing general trends in the learning process, while filtering out small oscillatory behaviour that could degrade the learning process. This allows a higher gain term η to be used thus speeding up the learning process.

2. **Cumulative update:** The interconnecting weights are only updated after several input-output pairs have been presented to the network, instead of after each training example. The errors, δ_j, are accumulated. The result of this is to try and filter out oscillatory effects that could be produced by updating of the weights after each training example.

3. **Error functions:** The error function that is used to define the error term δ in equation 10 is usually defined as being proportional to the square of the Euclidean distance between the actual output(o_k) and the desired output(d_k), $E = 0.5 * \sum(d_k - o_k)^2$. Cubic, $E = 0.33 * \sum(d_k - o_k)^3$, and quadratic, $E = 0.25 * \sum(d_k - o_k)^4$, values of the Euclidean distance can also be used to define the error term.

4. **Transfer functions:** In general, the nodes in a back-propagation network use the sigmoid as a transfer function to map the input of a node to its output. Another transfer function that is commonly used is the hyperbolic tangent, which is just a bipolar sigmoid and allows the output of a node to vary between -1 and 1.

5.2 Learning vector quantisation

Learning vector quantisation is a neural network that assigns input vectors to one of several classes. A self-organising map (SOM) is used to do the classification. This layer is trained to place different input vectors into different regions, each region representing a different class.

The basic LVQ network operates by assigning several nodes in the SOM to belong to each different class. A winning node c is defined as the node that has the shortest Euclidean distance from the input vector, $c = arg\ min\|x - m_i\|$. The weights of the inputs to each winning node are updated according to the following equations [13] :

If the winning node, c, is in the correct class then

$$m_c(t+1) = m_c(t) + \alpha(t)[x(t) - m_c(t) \tag{12a}$$

If the winning node, c, is in the incorrect class then

$$m_c(t+1) = m_c(t) - \alpha(t)[x(t) - m_c(t) \tag{12b}$$

5.2.1 Variations of the standard LVQ algorithm

Variations of the basic learning equations 12 can be implemented in order to try and speed up and improve the learning and classification processes.

1. **Conscience:** It may happen that one of the nodes in the SOM layer starts to win all the time. In order to combat this a conscience factor can be introduced to penalise a node that wins too often. This conscience factor is added to the distance calculated for each node.

2. **Repulsion:** In the early stages of learning it can help to turn off the repulsion term as shown in equation 12b. Thus vectors are only moved if the winning node is in the correct class. This helps training in cases where a vector has to move through a region that belongs to a different class.

3. **Refinement:** Once a network has been trained well, it may be necessary to try and resolve a few cases where incorrect classifications are taking place. The learning algorithm can be modified for those cases in which the winning node m_i is in the wrong class and the second best node m_j is in the correct class. The input vector must also fall within a predefined window that lies between

the two nodes. The node m_i is then moved away from the input vector and the node m_j is moved towards the input vector.

4. **Extended classification:** In order to provide better classification, instead of only using the winning node to define the class, all the nodes in a class are used. A Bayesian likelihood function is then computed for each class, with the winning class being the one that has the highest likelihood.

6 Results

Various combinations of training and test sets are used. Here we provide results for five situations, which are described in table 1. The training and test sets are created by manually classifying specific regions as belonging to one of two classes: nodules or background tissue.

Results are given for Haralick's features in table 2, and for Laws' features in table 3. Results are given for the test cases in table 1.

The learning rules, transfer functions, momentum and error functions of the BP networks are given in the tables. For Haralick's features, the BP networks have five nodes in the input layer, five nodes in the hidden layer and two output nodes. For Laws' features there are 25 input nodes, 5 nodes in the hidden layer and two output nodes. One hundred thousand training iterations are used.

Results are shown for an LVQ network without a conscience factor or refinement, using standard classification. Experiments with variations of the LVQ network did not improve the performance, therefore these results are not given.

7 Conclusions

In most cases the experiments were successful in identifying nodules for the limited problem of classifying regions as nodule or background tissue. The results indicate that the texture features are sufficiently consistent, within an x-ray and across different x-rays, to use a training set of x-rays for general nodule detection.

The following observations can be made from the results:

1. Haralick's features produced more consistent results than Laws' features. The performance could possibly be improved by using more of the features defined by Haralick.

2. Better overall results were obtained when using the images with Sobel enhanced texture.

3. Back propagation with a *Delta* or *Cumulative delta* learning rule, and a *sigmoid* transfer function produced the most consistent classification results and had lower training errors.

4. The results for test cases 1 and 2 indicate that the training sets for these cases were too small.

5. Test cases 3 and 4 illustrate that classification can be performed on different stages of lung cancer development. Test case 3 provided better results than 4 for Sobel images. This is probably due to the fact that the advanced cancer training set does not contain small nodules.

The procedures described in this paper could be used as a further classification stage after previous processing has identified possible nodule sites. A further investigation of texture analysis could possibly produce a system that could segment an image into regions of nodules and other anatomical classes.

Further research could involve using a feature vector that combines Haralick's and Laws' features. Only the best discriminating features should be selected and used for classification.

References

[1] M. L. Giger, K. Doi, and H. MacMahon, "Image feature analysis and computer-aided diagnosis in digital radiography. 3. automated detection of nodules in peripheral lung fields," *Medical Physics*, vol. 15, pp. 158–66, Mar/Apr 1988.

[2] D. H. Ballard and J. Sklansky, "A ladder-structured decision tree for recognizing tumors in chest radiographs," *IEEE Transactions on Computers*, vol. C-25, pp. 503–13, May 1976.

[3] S. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," *IEEE Transactions on Medical Imaging*, vol. 8, pp. 377–86, December 1989.

[4] M. L. Giger, N. Ahn, K. Doi, H. MacMahon, and C. E. Metz, "Computerized detection of pulmonary nodules in digital chest images: Use of morphological filters in reducing false-positive detections," *Medical Physics*, vol. 15, pp. 861–5, Sept/Oct 1990.

[5] W. Borchardt and G. de Jager, "The Hough transform as a circle finder," in *2nd South African Workshop on Pattern Recognition*, November 1991.

[6] W. Borchardt, "Computer-assisted detection of lung cancer nodules in medical chest x-rays," Master's thesis, University of Cape Town, September 1992.

[7] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, pp. 610–621, Nov. 1973.

[8] K. I. Laws, "Rapid texture identification," in *Proc. SPIE*, vol. 238, pp. 376–80, Dec. 1980.

[9] J. Y. Hsiao and A. A. Sawchuk, "Supervised textured image segmentation using feature smoothing and probabilistic relaxation techniques," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, pp. 1279–92, Dec. 1989.

[10] The Khoros Group, *Khoros user manual.* Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131.

[11] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, pp. 4–22, April 1987.

[12] NeuralWare, Inc, *Neural Computing.* User Manual for NeuralWorks Professional II/Plus.

[13] T. Kohonen, J. Kangas, J. Laaksonen, and K. Torkkola, *The Learning Vector Quantization Program Package version 2.1.* Helsinki University of Technology, Laboratory of Computer and Information Science, Rakentajanaukio 2 C, SF-02150 Espoo, Finland.

Table 1: Test cases

Test case	*Training set*	*Test set*
1	Right lung of advanced cancer	Left lung of advanced cancer
2	Right lung of intermediate cancer	Left lung of intermediate cancer
3	Right and left lung of intermediate cancer	Right and left lung of advanced cancer
4	Right and left lung of advanced cancer	Right and left lung of intermediate cancer
5	Right lung of advanced and right lung of intermediate cancer	Left lung of advanced and left lung of intermediate cancer

Table 2: Results using Haralick's features. Results are given as (% of nodules classified correctly)/(% of tissue classified correctly)

Parameters		*BP Networks*			*LVQ Network*
Learning		Norm-Cum-Delta	Delta-rule	Delta-rule	-
Transfer		Sigmoid	Sigmoid	TanH	-
Momentum		0.4	0.4	0.4	-
Error		Square	Square	Square	-
Test case	*Image type*				
1	Original	78/100	83/100	83/100	83/100
	Sobel	89/100	89/100	83/100	83/100
2	Original	80/67	80/67	80/67	97/75
	Sobel	90/50	90/50	90/50	83/89
3	Original	100/55	100/55	97/80	90/66
	Sobel	94/95	94/85	95/91	90/66
4	Original	80/94	89/94	92/87	94/94
	Sobel	92/87	92/87	96/81	89/94
5	Original	86/93	89/93	96/93	89/94
	Sobel	89/86	89/86	93/93	33/100

Table 3: Results using Laws' features. Results are given as (% of nodules classified correctly)/(% of tissue classified correctly)

Parameters		*BP Networks*			*LVQ Network*
Learning		Norm-Cum-Delta	Delta-rule	Delta-rule	-
Transfer		Sigmoid	Sigmoid	TanH	-
Momentum		0.4	0.4	0.4	-
Error		Square	Square	Square	-
Test case	*Image type*				
1	Original	78/100	83/100	89/100	89/100
	Sobel	83/100	83/100	83/100	87/100
2	Original	90/50	90/50	100/0	90/50
	Sobel	90/17	90/50	90/34	90/67
3	Original	75/95	81/100	69/90	41/100
	Sobel	92/90	95/90	95/95	84/89
4	Original	88/63	88/75	88/69	89/87
	Sobel	85/87	88/94	88/81	89/94
5	Original	85/100	80/100	83/100	86/100
	Sobel	82/100	82/100	82/100	74/100