SCALE-SPACE TEXTURE ANALYSIS[†]

*Andrew P. Bradley** *Paul T. Jackway** *Brian C. Lovell**

* *Dept. of Electrical and Computer Engineering, The University of Queensland, Brisbane, QLD 4072, AUSTRALIA.*

In this paper we propose a technique for classifying images by modeling features extracted at different scales. Specifically, we use texture measures derived from Pap smear cell nuclei images using a Grey Level Co-occurrence Matrix (GLCM). For a texture feature extracted from the GLCM at a number of distances we hypothesise that by modeling the feature as a *continuous* function of scale we can obtain information as to the shape of this function and hence improve its discriminatory power. This hypothesis is compared to the traditional method of selecting a given number of the best single distance measures. It is found, on the limited data set available, that the classification accuracy can be improved by modeling the texture features in this way.

1 INTRODUCTION

Scale is of vital importance in the analysis and understanding of signals. The adage "*You can't see the woods for the trees*" is a classic example of a problem of scale. A forest can only be recognised as such within a particular range of distances (scales). If you are too close, the forest appears as a single branch, piece of bark or collection of molecules. From a distance of hundreds of kilometers the forest just becomes a small part of the shape and texture of the landscape.

The idea that all the information in a signal is not contained at only one scale is of crucial importance. It has been shown that to fully analyse the structure of the signal it is necessary to relate information from a number of different scales. A method of combining this information, proposed by Witkin [15], is to treat scale as a *continuous* variable rather than a parameter. Signal features measured at different scales can then be related if they lie on the same feature path in "scale-space". In the computer vision community, a variety of image structures have been analysed at different scales by using this multi-scale representation [8]. A classic example of this scale-space analysis is in signal matching [14].

In this paper we have applied this same principle to the investigation of measures of image texture at varying scales, most measures of texture being determined at a number of scales or distances [3]. We propose a method to incorporate the information from all these scales in a meaningful way, hence significantly reducing the dimensionality of the data, whilst maintaining as much useful information as possible. The application we have chosen to illustrate this technique is cervical cell texture analysis using a Grey Level Co-occurrence Matrix (GLCM), therefore, some background to the GLCM will be given next.

2 GREY LEVEL CO-OCCURRENCE MATRIX TEXTURE MEASURES

The Grey Level Co-occurrence Matrix (GLCM) as proposed by Julesz [7] and later by Haralick *et al* [5], has been shown to be a powerful technique for measuring texture [1,4]. It is a second-order method that characterizes the probability that, given an image $f : D \subset \mathbb{Z}^2 \to [0, 1, 2, \ldots, n-1]$, the grey levels $k = f(i,j)$, and $l = f(i',j')$ co-occur. We define the distance between (i,j) and (i',j') as $\boldsymbol{d} = d\left((i,j),(i',j')\right)$, which expressed in polar co-ordinates is $r = |\boldsymbol{d}|$ and $\theta = \angle \boldsymbol{d}$.

The GLCM is then $\boldsymbol{C}_{r,\theta}$ where each element $c(k,l)$ is given by:

$$c(k,l) = Pr\left[f(i',j') = l \mid f(i,j) = k\right]. \tag{1}$$

The GLCM, $C_{r,\theta}$, is constructed by first quantising the image, f, into a "manageable" number of grey levels[1] $[0, 1, 2, \ldots, n-1]$, and then, for every pixel (i, j), examining the pixel (i', j') for specified values of r and θ. The GLCM, $C_{r,\theta}$, is then of size $n \times n$, with entries, $c(k, l)$, incremented every time the grey levels k and l co-occur. Probability estimates are obtained by dividing each entry in $C_{r,\theta}$ by the sum of all the entries. GLCMs are constructed at a number of distances, r, (scales) $[1, 2, 3, \ldots, m]$, and angles, θ, $[0°, 45°, 90°, 135°]$ in the image. **Note:** the GLCM is constructed to be symmetric so that $C_{r,\theta} = \frac{C_{r,\theta} + C_{r,\theta}^{\mathrm{T}}}{2}$, i.e., $C_{r,0°}^{\mathrm{T}} = C_{r,180°}$ $C_{r,45°}^{\mathrm{T}} = C_{r,225°}$, $C_{r,90°}^{\mathrm{T}} = C_{r,270°}$, and $C_{r,135°}^{\mathrm{T}} = C_{r,315°}$. Rotational invariance is then usually obtained by averaging with respect to θ [1].

Once the GLCM, $C_{r,\theta}$, has been constructed, its "content" is characterized using descriptors that extract features from $C_{r,\theta}$. For example, a descriptor that has a relatively high value when the values of $C_{r,\theta}$ are near the main diagonal, is the Inverse Difference Moment (Local Homogeneity):

$$IDM_{r,\theta} = \sum_k \sum_l \frac{c(k, l)}{(1 + (k - l)^2)}, \tag{2}$$

while a descriptor such as Entropy measures randomness, reaching its highest value when the elements of $C_{r,\theta}$ are equal:

$$Ent_{r,\theta} = -\sum_k \sum_l c(k, l) \ln c(k, l). \tag{3}$$

A number of other such features have been proposed [5]. The conventional method of texture analysis using the GLCM is to treat these features, extracted at different distances (scales), as independent features, selecting a small subset of scales (often a single scale) which gives the highest discriminatory power [16].
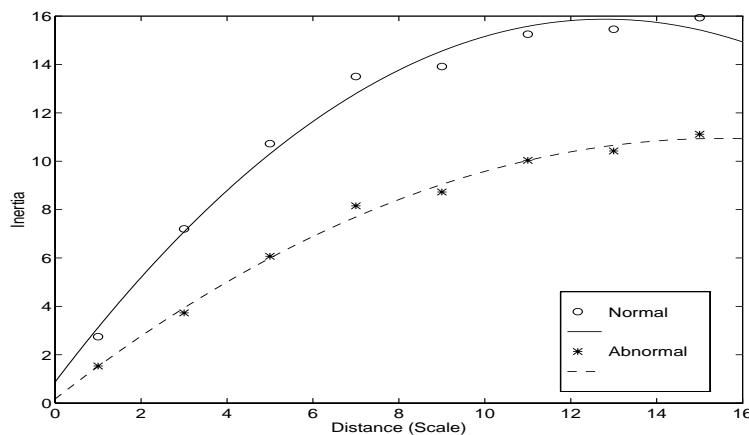

Figure 1: Sample points and their continuous scale-space texture functionals.

We propose to treat these measurements of texture at each scale as sample points of a continuous function through scale-space. This function can then be reconstructed by interpolating the sample points of the functions extracted from the GLCM, as shown in Figure 1. This allows a whole new range of classical mathematical techniques for comparing functions to be used in the analysis and classification of textures.

---

[1] In our case 16 [12].

3 TEXTURE AS A FUNCTION OF SCALE

Assume we have an image $f : D \subset \mathbb{R}^2 \to [0, 1, 2, \ldots, n-1]$, with the domain D bounded and of area

$$A(D) = \iint_D dx\,dy. \tag{4}$$

Since this image is defined on a continuous domain, we need to reformulate the theory of Grey Level Co-occurrence Matrices (GLCM) for the continuous case. To unify our work with scale-space theory, we will use the notation σ (scale) instead of r (distance) to denote pixel displacement. We define the Grey Level Co-occurrence Function (GLCF) $g_{\sigma,\theta} : [0, 1, 2, \ldots, n-1] \times [0, 1, 2, \ldots, n-1] \to \mathbb{R}$ as:

$$g_{\sigma,\theta}(\alpha, \beta) = \frac{1}{A(D)} \iint_D I\left[f(x, y) = \alpha,\ f(x', y') = \beta \right]\,dx\,dy, \tag{5}$$

where,

$$x' = x + \sigma \cos \theta, \tag{6}$$
$$y' = y + \sigma \sin \theta, \tag{7}$$

and the indicator function,

$$I\left[f(x, y) = \alpha,\ f(x', y') = \beta \right] = \begin{cases} 1 & \text{if } (f(x, y) = \alpha) \text{ and } (f(x', y') = \beta); \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

In other words, the GLCF estimates the probability that a pair of grey levels $[\alpha, \beta]$ will be found at a displacement $[\sigma \cos \theta, \sigma \sin \theta]$ apart.

Scale dependent texture features $T(\sigma)$ are extracted by applying some functional ψ to the GLCF. If this functional integrates $g_{\sigma\theta}(\alpha, \beta)$ over α, β and θ, then the texture feature of interest can be expressed as a function of scale. We write:

$$T(\sigma) = \psi\left[g_{\sigma,\theta}\left(\alpha, \beta\right) \right]. \tag{9}$$

The scale dependent functionals used in this work are continuous versions of those commonly used for discrete textural feature measures [1,5], for example, Energy:

$$En(\sigma) = \int_\theta \iint_D g_{\sigma,\theta}^2(\alpha, \beta)\,d\alpha\,d\beta\,d\theta \tag{10}$$

We also used: Inverse Difference Moment (Local Homogeneity); Entropy; Correlation; Inertia; Cluster Shade; and Cluster Prominence. In this paper we approximate these scale dependent functionals by fitting *continuous* functions to the discrete measurements extracted from the GLCM. However, we could have also fitted a continuous function to the image, allowing direct calculation of the GLCF. This would also allow scale to be extracted as a continuous variable.

4 TEXTURE ANALYSIS OF CERVICAL CELL NUCLEI

Traditionally, the computerized image analysis of cervical cells has attempted to use the same features to discriminate normal and abnormal cells as used by cytologists. These features, extracted from the nucleus and cytoplasm of each cell, are usually morphometric and photometric features, such as the size, shape, and optical density [9]. Additional clues such as the context of cells *i.e.*, free-lying or in a group, can then be used to improve classification performance. However, cell dysplasia (abnormality) is thought to initiate in the nucleus of the cell [10] and so features that measure the changing DNA structure in the nucleus have potential to be highly
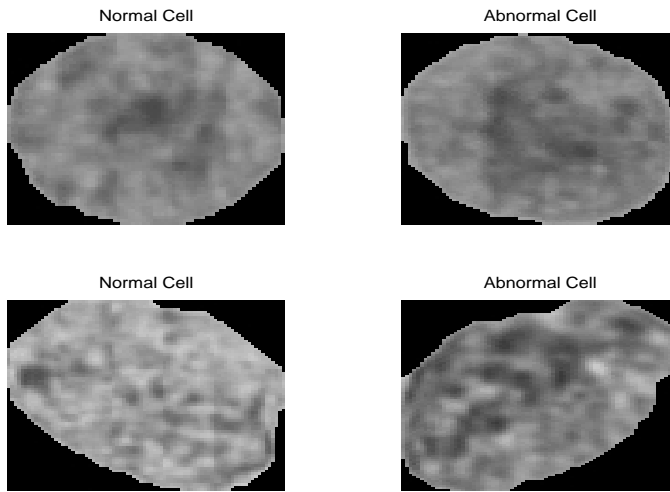
Figure 2: Examples of segmented normal and abnormal cervical cell nuclei.

discriminatory for the early stages of cell dysplasia. Features that measure this nuclear texture have been successfully used to discriminate between normal and abnormal cervical cells [12].

Figure 2 shows examples of the texture in typical normal and abnormal cervical cell nuclei. It can be appreciated from this figure that it is rather difficult for the untrained observer to distinguish between the two different cell types. In this paper the data set consisted of some 117 segmented cell nuclei, there being 58 normal and 59 abnormal cells. The grey-scale images were captured and segmented as detailed in [12]. For each cell a GLCM was created at distances of 1 to 15, at odd intervals only. Then the 7 texture features were extracted from each of the 8 GLCMs giving a total of 56 texture measurements for each cell. Continuous functions were then fitted to the 8 distance measurements for each of the texture functionals, as shown in Figure 1 for the Inertia functional. For the purpose of this study only polynomial models were considered, so that the 8 distance measurements were reduced to 2 (gradient and intercept) for a linear model , 3 for a quadratic, and 4 for a cubic. We then estimated the error rate of each model and selected the model with the lowest estimated error rate. Then, we selected the best N parameters from these best models and again classified the cervical cell data. This was compared to the conventional method of choosing N distances from the original 56 available[2].

In all cases features were first normalised to be in the range [0,1] and then selected using a Sequential Forward Search with the Bhattacharyya distance measure for determining feature separation [2,11]. The error estimate was obtained using leave-one-out [13] random sub-sampling using K-Nearest Neighbours (with $K = 3$) classification. This strategy is computationally intensive, but is generally considered to be one of the most reliable estimators of true error rate.

5   RESULTS

No. Features	1	2	3	4	5	6	7
Best Model	79%	85%	85%	84%	88%	88%	89%
Best Scales	83%	80%	85%	82%	87%	85%	86%

Table 1: Comparison of classification accuracies for the best N features.

---

[2]7 features each at 8 distances.

Table 1 shows estimates of the true accuracy for the conventional method, selecting N features from the original 56 distance measures, and for the proposed method, selecting N of the parameters from the best model for each feature.

6   DISCUSSION

The classification accuracy of the proposed technique was as good as, or better than, the convention technique in all cases except for classification using only a single feature. There is some evidence to suggest that the proposed method is better than the conventional one. However, on the current data set, this difference is not significant at the 0.10 level and so we are unable to reject the hypothesis that the two classification schemes produce equal accuracies[3]. Fitting models to each of the 7 texture features reduced the dimensionality of the original data set from 56 to 21. This is a significant reduction in dimensionality and reduces the computational complexity of further feature selection. Karhunen-Loéve analysis could also have been used to reduce the dimensionality, but in initial trials was found to perform poorly in this case, and, so far, has not been investigated further.

In general, the classification accuracy of each texture feature increased when we considered more than one scale. This confirms a result found by Conners and Harlow [1], and reiterates the fact that all the information in the signal is *not* contained at just one scale. The overall accuracy also increased when combining different texture features together, a similar result was found by Gotlieb and Kreyszig for their *atomic* and *composite* classifiers [4].

Scale should really be treated in a logarithmic manner, in this way changes in the texture function between, say scales 1 and 2 should be just as significant as changes in scale between 8 and 16. This was not done in this study as the data had already been collected at integer scales, but models should perhaps be fitted to a logarithmic scale in future research. Another advantage of the model fitting technique is that the features can be measured at as many distances or scales as you desire. The more samples of the *continuous* function the better, with the conventional method however, this would add to the dimensionality of the problem and make feature selection more computationally expensive. In this paper, only polynomial models have been discussed. There are, of course, a multitude of models and methods that could be used, such as autoregressive models, cubic splines, likelihood ratio test, *etc.* There are also a number of other scale dependent functionals that this technique can be applied to, for example features extracted from a morphological scale-space [6].

7   CONCLUSIONS

We have formulated a technique for classifying texture as a *continuous* function of scale. We have empirically shown the technique to perform better than the conventional method on textures derived from cervical cell nuclei. In addition, we have highlighted the following advantages of the proposed technique: It uses information from a number of scales by modeling the *shape* of the texture functions in scale-space; It reduces the dimensionality of the data, producing "high order" features; and Dimensionality is not increased by taking more distance measures.

8   ACKNOWLEDGEMENTS

---

[3]Using a paired, two tailed, t-test.

Walker who provided the original GLCM features. The software used for feature selection and error estimation was part of the **tooldiag** software package [2,11].

9 REFERENCES

[1] R. W. Conners and C. A. Harlow, "A theoretical comparison of texture algorithms," *IEEE Trans Pattern Anal Machine Intell*, PAMI-2, no. 3, pp. 204–222, 1980.

[2] P. A. Devijver and J. Kittler, in *Pattern Recognition: A statistical approach.* London: Prentice Hall, 1982.

[3] R. C. Gonzalez and R. E. Woods, *Digital Image processing.* Reading, Mass., Addison-Wesley, 1992.

[4] C. C. Gotlieb and H. E. Kreyszig, "Texture descriptors based on co-occurrence matrices," *Computer Vision, Graphics, and Image Processing*, 51, pp. 70–86, 1990.

[5] R. M. Haralick, K. S. Shanmugham and I. Dinstein, "Texture features for image classification," *IEEE Trans Syst Man Cybern*, SMC-3, no. 6, pp. 610–621, 1973.

[6] P. T. Jackway, "Morphological scale-space," in *Proceedings 11th IAPR International Conference on Pattern Recognition.* pp. 252–255, 1992.

[7] B. Julesz, "Visual pattern discrimination," *IRE Transactions on Information Theory*, Vol IT 8, pp. 84–92, 1962.

[8] T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales," *Journal of applied statistics*, Vol 21 , no. 2, pp. 225–270, 1994.

[9] R. J. Madachy and Y. S. Fu, "Image analysis for automatic classification of miotic cervical cells," in *Proceedings of the Annual International Conference of the IEEE Enginerring in Medicine and Biology Society.* New Orleans, LA: IEEE New York, USA, pp. 372–374, November 1988.

[10] K. J. Pienta, A. W. Partin and D. S. Goffey, "Cancer as a Disease of DNA Organization and Dynamic Structure," *Cancer Research*, 49, pp. 2525–2532, 1989.

[11] T. W. Rauber, M. M. Barata and A. S. Steiger-Garcao, "A Toolbox for the analysis and visualisation of sensor data in supervision," Universidade Nova de Lisboa, Portugal, Intelligent Robots group Technical report, 1993.

[12] R. Walker, P. Jackway, B. Lovell and D. Longstaff, "Classification of cervical cell nuclei using morphological segmentation and textural feature extraction," in *Proceedings of the 2nd Australian and New Zealand Conference on Intelligent Information Systems.* pp. 297–301, 1994.

[13] S. M. Weiss and I. Kapouleas, "An empirical comparison of pattern recognition, neural nets, and machine learning classification methods," in *Proceedings of the eleventh international joint conference on artificial intelligence.* San Mateo: Morgan Kaufmann, pp. 781–787, 1990.

[14] A. Witkin, D. Terzopoulos and M. Kass, "Signal Matching through Scale-space," The International Journal of Computer Vision, pp. 133–144, 1987.

[15] A. P. Witkin, "Scale-space filtering," in *Proceedings of the International Joint Conference on Artificial Intelligence.* Palo Alto: Kaufman, pp. 1019–1022, 1983.

[16] S. W. Zucker and D. Terzopoulos, "Finding structure in co-occurance matrices for texture analysis," *Computer graphics and image processing*, Vol 12, pp. 286–308, 1980.