

QUANTIFYING STRUCTURE REGULARITY IN FLUORESCENCE MICROSCOPY CELL IMAGES USING A NOVEL MULTI-DIMENSIONAL APPROXIMATE ENTROPY METRIC

Tom Marchant¹, Mark Murphy², Greg Madden², Christopher Moore¹

¹North Western Medical Physics, The Christie NHS Foundation Trust, Manchester, UK.

²General Engineering Research Institute, Liverpool John Moores University, Liverpool, UK.

ABSTRACT

Techniques such as fluorescence microscopy reveal in three dimensions the complex mechanical structure of the cell cytoskeleton, made up of actin filaments, microtubules and intermediate filaments. Methods to quantify the degree of order in this structure could prove useful in classifying cells and may be related to the structural integrity of the cell. Approximate Entropy (ApEn) is a regularity statistic that has previously been used to quantify randomness in 1D time series data. We have extended the ApEn concept to define a multi-dimensional parameter suitable for application to three dimensional image data. This 3D ApEn parameter was applied to fluorescence microscopy images showing actin filaments in normal and cancerous human prostate cells. Differences between the ApEn calculated from the different images were observed, with cancer cells tending to have higher values indicating a lower degree of order.

Index Terms— fluorescence microscopy, approximate entropy, prostate cancer, image texture, regularity

1. INTRODUCTION

The cytoskeleton is a highly complex, dynamic three-dimensional structure, which can be imaged using techniques such as laser scanning fluorescence microscopy. Often large volumes of data are produced, requiring automated analysis to extract meaningful information, leading to the emerging field of bioimage informatics [1]. Key tasks in interpreting cell images include feature extraction, segmentation, and classification [2]. The actin cytoskeleton is instrumental in determining a cell's shape [3]. Cancer is strongly associated with disordered changes in the cytoskeletal structure [4]. Hence it may be possible to recognize early malignancy by developing tools to analyze cytoskeletal image irregularity rather than waiting for obvious changes in cell shape/morphology. This would address the key challenge of subtle progression of malignancy between normal and diseased structural states, which currently requires expert histopathological interpretation to grade in terms of cell 'differentiation'. In turn, this would inform automatic classification of cells as well as modeling of cytoskeleton structure and mechanical strength.

Approximate Entropy (ApEn) is a regularity statistic in which the order of data and, consequently, inherent pattern is important [5]. This quality allows it to provide important additional information compared to variability statistics (exemplified by mean, standard deviation, etc.) in the analysis of medical time series data such as fetal heart rate variability [6], particularly the ability to aid classification between different groups of patients. However, ApEn has found limited application to image data analysis, since it has only been defined for 1D data sequences.

Following previous work using approximate entropy to classify perturbations in the speech patterns of larynx cancer patients [7], North Western Medical Physics has developed an extension to the original ApEn concept which allows it to be calculated for 2D and 3D image data, with clinical applications in mind. Here we describe the procedure for calculating the new multidimensional ApEn parameter and illustrate its application to the analysis of fluorescence microscopy images of normal and cancerous human prostate cells from Liverpool John Moores University.

This paper was presented at IEEE International Conference on Image Processing, Brussels, Belgium, September 2011.

<http://dx.doi.org/10.1109/ICIP.2011.6116317>

© 2011 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

2. METHODS & MATERIALS

2.1. Approximate Entropy

Approximate Entropy [5] measures the probability that two segments of data within a time series that are similar (within some threshold) will remain similar when the length of both segments is incremented by one. Three parameters, m , r and N , are required to define ApEn: m indicates the length of segments within the data to be compared; r is the threshold parameter for judging similarity of data segments; and N is the overall length of the sequence for which ApEn is to be calculated. ApEn is then defined as

$$\text{ApEn} = \ln \left[\frac{C_m(r)}{C_{m+1}(r)} \right], \quad (1)$$

where $C_m(r)$ represents the prevalence of similar patterns of length m in the data. This is calculated by considering each segment of length m in turn and finding the fraction of all length- m segments that are similar to it. Averaging this fraction over all segments yields the value of $C_m(r)$. Similarity between two segments is defined by calculating the maximum absolute difference between any element in one segment and the corresponding element in the other segment, and comparing this to the threshold parameter, r .

Formally, from N data points $\{u(i)\} = u(1), u(2), \dots, u(N)$, a series of vector sequences $\mathbf{x}(1)$ to $\mathbf{x}(N - m + 1)$ are formed composed of m consecutive u values: $\mathbf{x}(i) = [u(i), \dots, u(i + m - 1)]$. The sequence $\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N - m + 1)$ is used to construct $C_m^i(r)$ for each i , $1 \leq i \leq (N - m + 1)$,

$C_m^i(r) = (\text{number of } j \text{ such that } d[\mathbf{x}(i), \mathbf{x}(j)] \leq r) / (N - m + 1)$. Here $d[\mathbf{x}(i), \mathbf{x}(j)]$ is the “distance” between vectors defined as the maximum difference in their corresponding scalar components. Averaging $C_m^i(r)$ over i gives

$C_m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} C_m^i(r)$, and ApEn is calculated according to equation (1) above.

It is convenient to define the threshold parameter, r , as proportional to the standard deviation of the data, $r = k \times SD$, where k is a constant. Lower ApEn value indicates the presence of repeating patterns in the data and hence a higher degree of order. With more randomness the probability that incremented segments remain similar is reduced, resulting in higher ApEn.

2.2 Calculation of ApEn for 3D image data

When defining an ApEn parameter for use in three dimensions, the aim is to calculate a local ApEn value based on the data values in a region of interest surrounding the point under consideration. This can be the immediate $3 \times 3 \times 3$ cube surrounding that point, or a larger region (e.g. $5 \times 5 \times 5$, $7 \times 7 \times 7$, etc.). It is convenient to define N as the side length of this cube and hence the number of data points over which ApEn will be calculated is N^3 .

For application to 3D data, one must decide how to interpret a sequence of data starting at a particular point (the $\mathbf{x}(i)$ vectors). Three possibilities are (1) to extract a 1D data sequence in a pre-defined direction (as in [8]), (2) to extract 1D data in a raster fashion (as in [9,10]), or (3) to extend the sequence outwards from the initial point in all directions, so that the 26 nearest neighbor points surrounding the initial point are collectively considered as the second ‘element’ of $\mathbf{x}(i)$, the 98 next nearest neighbors form the third ‘element’ of $\mathbf{x}(i)$ and so on. We have chosen the last of these options as it removes directional dependence from the ApEn_{3D} parameter, thus avoiding assumptions about the orientation of features of interest.

Defining $\mathbf{x}(i)$ in this way, so that its elements are no longer all scalar values, requires a modified method to calculate $d[\mathbf{x}(i), \mathbf{x}(j)]$, the “distance” between two vectors. This requires a difference between each element of $\mathbf{x}(i)$ and the corresponding element of $\mathbf{x}(j)$ to be calculated. For elements consisting of multiple values (i.e. all except the first element) the difference could be defined as the maximum difference between any of the corresponding values. This would be maximally consistent with 1D ApEn calculation, but would require all voxels to match the compared sequence, within tolerance r . This rather strict criterion would result in a small number of sequences

classified as similar (thus increasing noise in the calculated ApEn values). A more pragmatic approach is to define the difference between vector elements of \mathbf{x} as the mean of the absolute differences between the corresponding values. Thus a series of scalar values, $C_m^i(r)$, are generated, from which ApEn_{3D} can be calculated.

It is important that our modifications to the ApEn algorithm for application to 3D data should not compromise the ability of ApEn to distinguish random from regular patterns. Validation tests were carried out to show this, firstly using a similarly modified 1D ApEn parameter and secondly by applying the new ApEn_{3D} parameter to artificially generated regular and random 3D test patterns. These tests demonstrated that our modified calculation methodology for ApEn behaved very similarly to conventional 1D ApEn and was able to distinguish regular from random 3D test patterns (see figure 1).

In practice we limit the calculation of ApEn_{3D} to $m=1$ for 3D data, as the calculation time increases rapidly for larger values of m . Also it is important that sufficient numbers of point and extended segment matches should be produced for a statistically useful estimate of ApEn to be calculated. For larger values of m the number of matches is decreased, meaning that larger values of N are required. Pincus estimated that for 1D data N should be at least 10^m [6]. We typically calculate ApEn_{3D} with $m=1$ and $N=3$ (i.e. 27 samples compared), which satisfies this requirement.

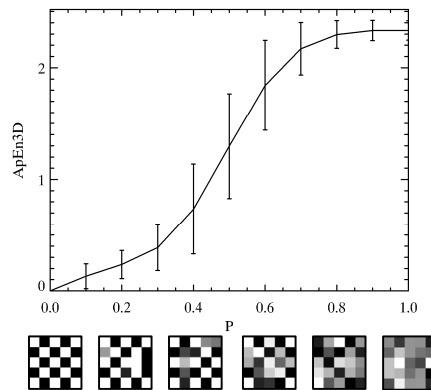


Fig. 1. ApEn_{3D} for regular 3D checkerboard pattern with increasing fraction (P) of points chosen at random replaced with uniformly distributed random noise. Each point shows the mean and standard deviation of 1000 randomly generated patterns for that noise fraction, P . Example 2D sections from the 3D test patterns are shown below the x-axis for a range of P values.

2.3. Cell image data

Fluorescence microscopy images of two cell lines were used in this work: prostate normal tissue (PNT) and prostate adenocarcinoma (PC-3). Cells of each type were grown in a monolayer, fixed and the actin filaments labeled using phalloidin conjugated with the fluorescent molecule FITC (Sigma, UK). Image stacks were acquired with a Zeiss LSM 510 laser scanning fluorescent microscope and using an excitation wavelength of 488 nm. Image size (x/y/z) was 512/512/98 for PNT and 512/512/81 for PC-3, with in-plane voxel size of 0.2 μm and slice (z) thickness 0.1 μm .

2.4. Image comparisons

For both the prostate normal tissue (PNT) and prostate cancer (PC-3) images the equator slice through the monolayer was selected (i.e. the z-slice showing the widest point on the cells, approximately half way from base to apex). ApEn_{3D} ($m=1$, $N=3$, $k=0.8$) was calculated for each voxel of the equator slices (software implemented using IDL, ITT VIS, Boulder, CO). Regions of interest (ROIs) were defined

manually covering the cell contents of 6-7 typical cells on each image. In particular the cytoskeleton, located within cytosol, was selected. Histograms of the $ApEn_{3D}$ values within each individual cell ROI and the combined ROI were generated and the mean values calculated. Histograms and mean values were compared between PNT and PC-3 cells. For comparison to $ApEn_{3D}$ the Sobel approximation of image gradient magnitude was also calculated for both images. Histograms of image gradient magnitudes within the cell ROIs were also generated and compared in the same way. This was designed to demonstrate the calibrated, structural discriminatory power of the ApEn regularity statistic compared to the more common forms of grey-scale image processing.

3. RESULTS

Figures 2a and 2b show the selected slices for PNT and PC-3 cells, showing the equator slice through the monolayer. Figures 2c and 2d show the same slices after $ApEn_{3D}$ processing. Computation of $ApEn_{3D}$ ($m=1$, $N=3$, $k=0.8$) for a 512^2 image slice takes approximately 4 minutes using a 3GHz CPU with 2GB RAM.

Regions containing little or no regular structure (i.e. flat or dominated by random noise) have high values of $ApEn_{3D}$, while areas displaying regularity (e.g. edges, regular textures) are characterized by low values of $ApEn_{3D}$. Note that an exactly flat image area would have low (zero) $ApEn_{3D}$ as this would be highly regular. However due to even small amounts of noise inherent in the imaging process, and the fact that the comparison threshold is defined as a fraction of the data standard deviation in the analysis region, flat regions of the image appear as pure noise (high $ApEn_{3D}$). Figures 2e and 2f show the corresponding Sobel gradient magnitude images.

Figures 3a and 3b show the cell ROIs defined on PNT and PC-3 cell images. Figures 3c and 3d show histograms of the $ApEn_{3D}$ values within the cell ROIs (black line shows histogram for the combined ROI, grey lines show histograms for individual cell ROIs). The histograms can be interpreted as containing a peak at high $ApEn_{3D}$ (around 2.2), corresponding to image areas with little structural regularity, along with a tail at lower $ApEn_{3D}$ values, corresponding to image regions containing regular structure. The histogram for the PNT cells shows a smaller peak at high values and a much more prominent tail at lower $ApEn_{3D}$ values, indicating that more regular structure is apparent in this image than the PC-3 image. The histograms for individual cells show that while there is some degree of variability between cells within each sample, the difference in histogram shape between the two samples remains consistent. The combined ROI histograms for PNT and PC-3 cells are compared directly in figure 4a. Both histograms are normalized to an area of 1.0, hence their similarity of shape can be quantified by the common area under both curves, 0.63.

These differences in $ApEn_{3D}$ histograms are also reflected in the mean $ApEn_{3D}$ values. The PNT combined cell ROI has a mean $ApEn_{3D}$ of 1.06 ± 0.12 (SD of mean values for individual cell ROIs), while the PC-3 cells have mean $ApEn_{3D}$ of 1.59 ± 0.06 .

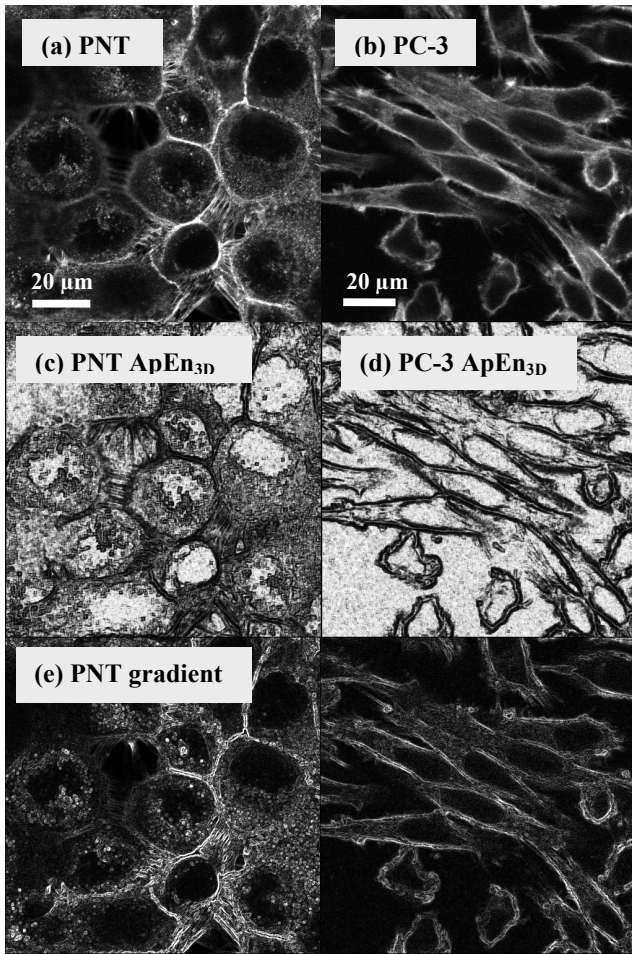


Fig. 2. (a) and (b) Equator slices through actin labeled fluorescence microscopy images of PNT and PC-3 cells. (c) and (d) $ApEn_{3D}$ processed images corresponding to the slices shown in (a) and (b). ($ApEn_{3D}$ images are scaled to show values in the range 0.0-2.5). (e) and (f) Image gradient magnitude (Sobel) processed images for PNT and PC-3 cells (gradient images are scaled to show values in the range 0-600).

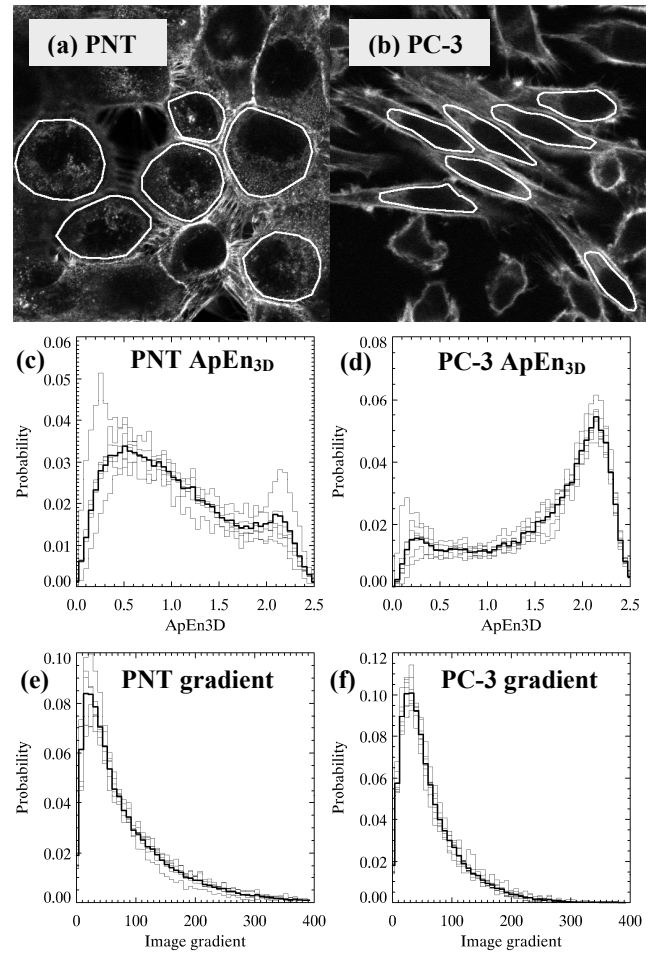


Fig. 3. (a) Cell ROIs defined on PNT image (b) Cell ROIs defined on PC-3 image (c) Histogram of $ApEn_{3D}$ values within PNT cell ROIs (grey lines show histograms for individual cells, black line shows combined ROI histogram) (d) Histogram of $ApEn_{3D}$ values within PC-3 cell ROIs (e) Histogram of image gradient values within PNT cell ROIs (f) Histogram of image gradient values within PC-3 cell ROIs.

Figures 3e, 3f and 4b show the corresponding gradient magnitude distributions for PNT and PC-3 cells for comparison. The differentiation between the two cell types is less significant. The combined ROI gradient histograms shown in figure 4b have an overlap area of 0.88.

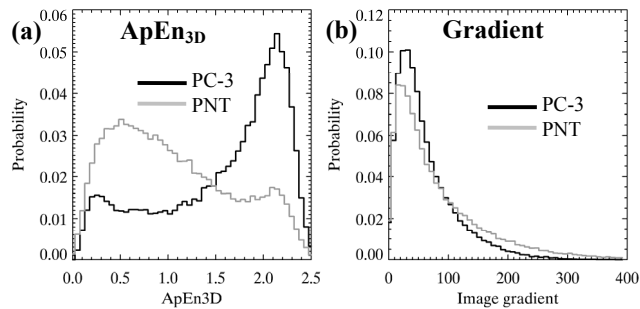


Fig. 4. (a) Combined cell ROIs $ApEn_{3D}$ histogram for PNT (grey) and PC-3 (black) cells (b) Combined cell ROIs image gradient histogram for PNT (grey) and PC-3 (black) cells.

4. DISCUSSION

We have described the application of a new multi-dimensional extension of the regularity statistic Approximate Entropy to fluorescence microscopy image data. Our focus was on calibrated differentiation of the actin cytoskeleton structure of normal and cancerous human prostate cells. The cancer cells indeed showed a higher mean $ApEn_{3D}$ value than the normal cells, indicating a greater degree of randomness in the image data. This is in agreement with the observation that the actin filaments of some cancer cells are disrupted causing the cells to have increased deformability compared to healthy cells [4]. Recent studies of the effect of radiation on the cytoskeleton have also observed changes in the organization of actin filaments [11]. Hence methods to quantify image regularity may also be of use here.

An important feature of ApEn is that it is a statistical measure sensitive to the range of inherent patterns and structures present in the image data itself, whose recurrence and extent are an integral part of the computational process. Hence, $ApEn_{3D}$ also responds to simplistic features such as edges, which can be visualised using gradient processing. This study shows that ApEn has distinctive, additional value in being able to clearly differentiate between regular and irregular image patterns at a more subtle level, as demonstrated by the comparison to gradient processed cell images.

The computation of ApEn in 3D rather than 2D includes additional data from adjacent image slices. This increases the number of data points available (thus decreasing noise) and takes account of regularity in all available planes, which may not be evident from a single slice extracted from the data volume.

In this work we used manual outlining of cell contents, however automatic detection of cell boundaries (e.g. [12]) could be implemented to make analysis of larger samples of images feasible.

5. CONCLUSION

Three dimensional approximate entropy processing of cell images may be a useful tool for measuring structure regularity of the cytoskeleton, which could help in classification of cells and modeling of cell structure.

6. REFERENCES

- [1] H. Peng, "Bioimage informatics: a new area of engineering biology," *Bioinformatics*, 24 (17) pp. 1827-1836, 2008.
- [2] R. Murphy, M. Velliste and G. Porreca, "Robust numerical features for description and classification of subcellular location patterns in fluorescence microscope images," *Journal of VLSI Signal Processing*, 35 (3) pp. 311-321, 2003.
- [3] Alberts B, et al., *Molecular Biology of the Cell*, 4th Edition, Garland Science, New York, 2002.

- [4] Kellie S, "The cytoskeleton and neoplastic transformation", in *The Cytoskeleton: A Multi-Volume Treatise*, 3, pp. 133-158, 1996.
- [5] S. Pincus, "Approximate Entropy as a measure of system complexity," *Proc Natl Acad Sci USA*, 88 pp. 2297-2301, 1991.
- [6] S. Pincus and A. Goldberger, "Physiological time-series analysis: what does regularity quantify?" *AJP - Heart and Circulatory Physiology*, 266 pp. H1643-1656, 1994.
- [7] C. Moore, K. Manickam and N. Slevin, "Collective spectral pattern complexity analysis of voicing in normal males and larynx cancer patients following radiotherapy," *Biomedical Signal Processing and Control*, 1 (2) pp. 113-119, 2006.
- [8] B. Singer and S. Pincus, "Irregular arrays and Randomization," *Proc Natl Acad Sci USA*, 95 (4) pp. 1363-1368, 1998.
- [9] G. Parker, J. Schnabel, et al., "Nonlinear smoothing of MR images using Approximate Entropy — a local measure of signal intensity irregularity," *Information Processing in Medical Imaging '99, LCNS 1613* pp. 484-489, 1999.
- [10] J. Cullen, A. Saleem, et al., "Measurement of cardiac synchrony using Approximate Entropy applied to nuclear medicine scans," *Biomedical Signal Processing and Control*, 5 (1) pp. 32-36, 2010.
- [11] D. Gabryś, O. Greco, et al., "Radiation effects on the cytoskeleton of endothelial cells and endothelial monolayer permeability," *International Journal of Radiation Oncology*Biology*Physics*, 69 (5) pp. 1553-1562, 2007.
- [12] T. Jones, A. Carpenter and P. Golland, "Voronoi-based segmentation of cells on image manifolds," *Computer Vision for Biomedical Image Applications, LCNS 3765* pp. 535-543, 2005.

ACKNOWLEDGEMENT

This work was funded in part by the Engineering and Physical Sciences Research Council (UK) under grant EP/H024913/1.