

Cell Orientation Entropy (CORe): Predicting Biochemical Recurrence from Prostate Cancer Tissue Microarrays

George Lee¹, Sahirzeeshan Ali², Robert Veltri³, Jonathan I. Epstein³,
Christhunesa Christudass³, and Anant Madabhushi^{2,*}

¹ Rutgers, The State University of New Jersey. Piscataway, NJ, USA

² Case Western Reserve University, Cleveland, OH, USA

³ The Johns Hopkins Hospital, Baltimore, MD, USA

Abstract. We introduce a novel feature descriptor to describe cancer cells called Cell Orientation Entropy (CORe). The main objective of this work is to employ CORe to quantitatively model disorder of cell/nuclear orientation within local neighborhoods and evaluate whether these measurements of directional disorder are correlated with biochemical recurrence (BCR) in prostate cancer (CaP) patients. CORe has a number of novel attributes that are unique to digital pathology image analysis. Firstly, it is the first rigorous attempt to quantitatively model cell/nuclear orientation. Secondly, it provides for modeling of local cell networks via construction of subgraphs. Thirdly, it allows for quantifying the disorder in local cell orientation via second order statistical features. We evaluated the ability of 39 CORe features to capture the characteristics of cell orientation in CaP tissue microarray (TMA) images in order to predict 10 year BCR in men with CaP following radical prostatectomy. Randomized 3-fold cross-validation via a random forest classifier evaluated on a combination of CORe and other nuclear features achieved an accuracy of $82.7 \pm 3.1\%$ on a dataset of 19 BCR and 20 non-recurrence patients. Our results suggest that CORe features could be extended to characterize disease states in other histological cancer images in addition to prostate cancer.

1 Introduction

In this paper, we developed a new approach to quantitatively characterize prostate cancer (CaP) morphology via cell orientation entropy (CORe) and thereby attempt to predict biochemical recurrence (BCR), a strong marker for presence of recurring cancer following radical prostatectomy (RP) treatment. BCR is defined by a detectable persistence of prostate specific antigen (PSA) of 0.2 ng/mL following RP. Nearly 60,000 patients undergo RP treatment for CaP each year, and

* Research reported in this publication was supported by the Department of Defense W81XWH-12-1-0171 and the National Cancer Institute of the National Institutes of Health under award numbers R01CA136535-01, R01CA140772-01, R43EB015199-01, and R03CA143991-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

for 15-40% of RP patients, BCR occurs within 5 years [1]. Gleason scoring (GS) is a qualitative system (2-10) which uses gland morphology to grade CaP aggressiveness and is representative of the clinical standard for predicting BCR. High GS 8-10 cases have been found to be correlated with BCR and presence of aggressive disease and often secondary treatment is provided to accompany RP based on the identification of high GS. Meanwhile, patients with GS 6 typically have a very low incidence of BCR and would not indicate a need for secondary treatment. Unfortunately, outcomes of intermediate GS 7 cancers can vary considerably, and statistical tables suggest a 5-year BCR-free survival rate as low as 43% in these men [2]. As such, predicting BCR in GS 7 cases is an important and largely unsolved problem with significant clinical and therapeutic implications.

While pathologists have traditionally used microscopic evaluation of histological tissue to determine the extent and severity of cancer, the recent advent of digital whole slide scanners has allowed for the development of quantitative histomorphometry (QH) for automated evaluation of histological tissue. The main idea behind these QH methods is to model the appearance of tumor morphology on histopathology via shape, textural, and spatio-architectural descriptors. While qualitative cancer grading remains by far the single most important prognostic measure of aggressive disease, it is subjective and prone to inter-reviewer variability among pathologists [3].

Many researchers have attempted to develop automated, computerized grading algorithms to address the problems of inter-reviewer variability in cancer grading and thereby improve classification accuracy [4,5,6,7,8]. Jafari-Khouzani et al. [6] examined the role of image texture features based on co-occurrence matrices for the purpose of automated CaP grading. However, these matrices are based on pixel intensity and lack direct biological significance. Tabesh et al. [5] also looked at color, texture, and structural morphology to evaluate prostate histopathology in terms of grading. However, complex spatial relationships between structures are not investigated.

Graph tessellations of cell nuclei using Voronoi or Delaunay graphs aim to describe the spatial interactions between nuclei in the tissue and have previously been found to be predictive of CaP grade [4]. However, these features are derived from fully connected graphs, whose edges traverse across epithelial and stromal regions. By connecting globally, fully connected graphs tend to dilute the contribution of the tumor morphologic features specific to the cancer epithelium. Therefore, global graphs are not sensitive to local cell organization, which may be critical in characterizing tumor aggressiveness.

Analysis of local subgraphs, which unlike global graphs (e.g. Voronoi and Delaunay) that aim to capture a global architectural signature for the tumor, can allow for quantification of local interactions within flexible localized neighborhoods. Bilgen et al. [7] constructed different types of cell graphs for evaluating breast cancer. In [8], Veltri et al. investigated nuclear morphology using a descriptor called nuclear roundness variance. Cell morphology was found to exceed Gleason scoring for predicting CaP aggressiveness.

In this paper, we present a new set of QH features, cell orientation entropy (CORe), which aim to capture the local directional information of epithelial cancer cells. CaP is fundamentally a disease of glandular disorganization and the resulting breakdown in nuclei orientation is related to its grade [9]. Epithelial cells align themselves with respect to the glands, and thus display a coherent directionality. However, cancerous prostate glands are less well formed, resulting in a more chaotic organization and orientation of the surrounding nuclei.

CORe attempts to model this difference between cancerous and benign regions via a novel scheme, unique to digital pathology image analysis. Firstly, it is the first rigorous attempt to quantitatively model cell orientation and explore the linkage between cell orientation and CaP aggressiveness. Secondly, while previous work has focused on global graph networks for characterizing tumor architecture, CORe employs subgraphs to construct local cell networks and thereby quantify second order statistics based on co-occurrence matrices of cell orientations. While co-occurrence matrices are commonly used to describe image textures [10], by quantifying second order statistics of image intensities, this is the first instance of the use of the co-occurrence matrix to evaluate local, higher order interactions of nuclear orientations. These second order local statistical features of nuclear orientation yield a rich set of descriptors for distinguishing the different CaP tumor classes.

2 Cell Orientation Entropy (CORe)

2.1 Automated Cell Segmentation

We employed an energy based segmentation scheme presented in [11] to detect and segment a set of cell/nuclei $\gamma_i, p \in \{1, 2, \dots, n\}$, where n is the total number of nuclei found. This segmentation scheme is a synergy of boundary and region-based active contour models that incorporates shape priors in a level set formulation with automated initialization based on watershed. The energy functional of the active contour is comprised of three terms. The combined shape, boundary and region-based functional formulation [11] is given below:

$$F = \underbrace{\beta_s \int_{\Omega} (\phi(\mathbf{x}) - \psi(\mathbf{x}))^2 |\nabla \phi| \delta(\phi) d\mathbf{x}}_{\text{Shape+boundary force}} + \underbrace{\beta_r \int_{\Omega} \Theta_{in} H_{\psi} d\mathbf{x} + \int_{\Omega} \Theta_{out} H_{-\psi} d\mathbf{x}}_{\text{Region force}} \quad (1)$$

where $\beta_s, \beta_r > 0$ are constants that balance contributions of the boundary based shape prior and the region term. $\{\phi\}$ is a level set function, ψ is the shape prior, $\delta(\phi)$ is the contour measure on $\{\phi = 0\}$, $H(\cdot)$ is the Heaviside function, $\Theta_r = |I - u_r|^2 + \mu |\nabla u_r|^2$ and $r \in \{in, out\}$.

The first term is the prior shape term modeled on the prostate nuclei, thereby constraining the deformation achievable by the active contour. The second term, a boundary-based term detects the nuclear boundaries from image gradients. The third term drives the shape prior and the contour towards the nuclear boundary based on region statistics.

2.2 Calculating Cell Orientation

To determine the directionality for each cell γ_i , we perform principal component analysis on a set of boundary points $[x_i, y_i]$ to obtain the principal components $Z = [z_1, z_2]$. The first principal component z_1 describes the directionality of the cell in the form of the major axis $z_1 = \langle z_1^x, z_1^y \rangle$, along which the greatest variance occurs in the nuclear boundary. The principal axis z_1 is converted to an angle $\bar{\theta}(\gamma_i) \in [0^\circ 180^\circ]$ counterclockwise from the vector $\langle 1, 0 \rangle$ by $\bar{\theta}(\gamma_i) = \frac{180^\circ}{\pi} \arctan(\frac{z_1^y}{z_1^x})$.

2.3 Local Cell Subgraphs

Pairwise spatial relationships between cells are defined via sparsified graphs. A graph $G = \{V, E\}$, where V represents the set of n nuclear centroids $\gamma_i, \gamma_j \in V$, $i, j \in \{1, 2, \dots, n\}$ as nodes, and E represents the set of edges which connect them. The edges between all pairs of nodes γ_i, γ_j are determined via the probabilistic decaying function

$$E = \{(i, j) : r < d(i, j)^{-\alpha}, \forall \gamma_i, \gamma_j \in V\}, \quad (2)$$

where $d(i, j)$ represents the Euclidean distance between γ_i and γ_j . $\alpha \geq 0$ controls the density of the graph, where α approaching 0 represents a high probability of connecting nodes while α approaching ∞ represents a low probability. $r \in [0, 1]$ is an empirically determined edge threshold.

2.4 Calculating Second Order Statistics for Cell Orientation

The objects of interest for calculating CORe features are the cell directions given by a discretization of the angles $\bar{\theta}(\gamma_i)$, such that $\theta(\gamma_i) = \omega \times \text{ceil}(\frac{\bar{\theta}}{\omega})$, where ω is a discretization factor. Neighbors defined by the local cell subgraphs G , allow us to define neighborhoods for each cell. For each $\gamma_i \in V$, we define a neighborhood $\mathcal{N}_i$, to include all $\gamma_j \in V$ where a path between γ_i and γ_j exists in graph G .

An $N \times N$ co-occurrence matrix $\mathcal{C}$ subsequently captures angle pairs which co-occur in each neighborhood $\mathcal{N}_i$, such that for each $\mathcal{N}_i$,

$$\mathcal{C}_{\mathcal{N}_i}(a, b) = \sum_{\gamma_i, \gamma_j}^{\mathcal{N}_i} \sum_{a, b=1}^N \begin{cases} 1, & \text{if } \theta(\gamma_i)=a \text{ and } \theta(\gamma_j)=b \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where $N = \frac{180}{\omega}$, the number of discrete angular bins. We then extract second order statistical features (Contrast energy, Contrast inverse moment, Contrast average, Contrast variance, Contrast entropy, Intensity average, Intensity variance, Intensity entropy, Entropy, Energy, Correlation, Information measure 1, Information measure 2) from each co-occurrence matrix $\mathcal{C}_{\mathcal{N}_i}(a, b)$. Selected formulations are described in Table 1. Mean, standard deviation, and range of Θ across all $\mathcal{N}_i$ constitute the set of 39 CORe features.

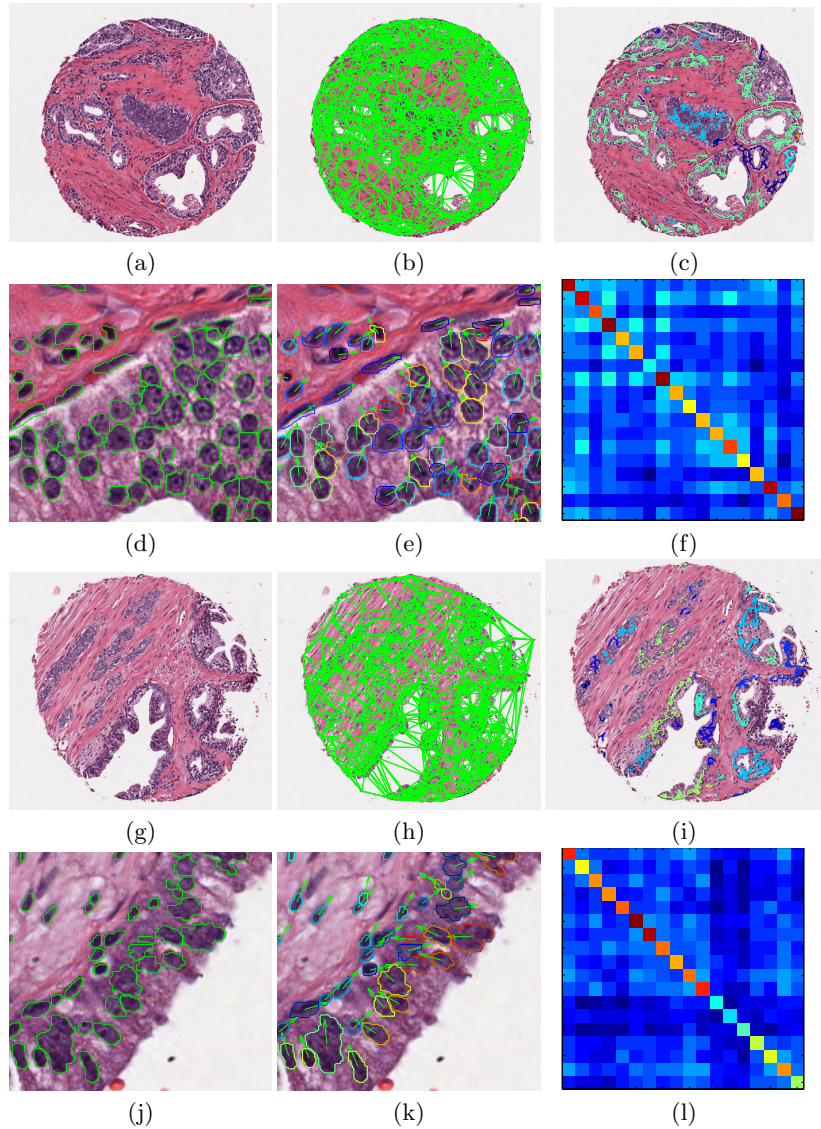


Fig. 1. Prostate TMAs pertaining to (a)-(f) BCR and (g)-(l) NR case studies. Nuclei are used as nodes for calculation of (b),(h) Delaunay graphs. Automated segmentation (d),(j) defines the nuclear boundaries and locations from the TMA image. (e),(k) Cell orientation vectors are calculated from the segmented boundaries (illustrated via different boundary colors). (c),(i) Subgraphs are formed by connecting neighboring cells. CORE features calculate contrast in the cell orientation (with dark regions showing more angular coherence and bright regions showing more disorder). Summation of the co-occurrence matrices provide a visual interpretation of disorder, where (f) shows brighter co-occurrence values in the off-diagonal cells, suggesting higher co-occurrence of nuclei of differing orientations compared to (l).

Table 1. Representative CORe features

CORe Feature (θ)	Description
Entropy	$\sum_{a,b} -\mathcal{C}(a,b) \log(\mathcal{C}(a,b))$
Energy	$\sum_{a,b} \mathcal{C}(a,b)^2$
Correlation	$\sum_{a,b} \frac{(a-\mu_a)(b-\mu_b)\mathcal{C}(a,b)}{\sigma_a\sigma_b}$
Contrast (variance)	$\sum_{a,b} a-b ^2 \mathcal{C}(a,b)$

3 Experimental Design

3.1 Prostate Cancer Tissue Microarray Data

While CORe is extensible towards the histological analysis of other pathological diseases, we have chosen prostate cancer (CaP) as a test case for this initial work. Our dataset comprised of histologic image samples in the form of tissue microarray (TMA) cores from 19 CaP patients who experienced BCR within 10 years of RP, and from 20 patients who did not (NR). Patients were matched for GS 7 and tumor stage 3A. CaP tissue included in the TMAs were selected and reviewed by an expert pathologist. For this study, each of 39 patients was represented by a single randomly selected 0.6mm TMA core image, chosen from a set of 4 TMA cores taken for that patient.

3.2 Comparative Methods for Evaluating CORe

We compared the efficacy of CORe features with previously studied nuclear features. The shape of individual nuclei has previously been shown to be prognostic of GS [8,12]. The set of 100 cell morphology features representing mean, standard deviation of nuclear size and shape are summarized in Table 2.

Nuclear/cell architecture refers to the spatial arrangement of cells in cancerous and benign tissue. 51 architectural image features describing the nuclear arrangement were extracted as described in [12]. Voronoi diagrams, Delaunay Triangulation and Minimum Spanning Trees were constructed on the digital histologic image using the nuclear centroids as vertices (See Table 2).

For all feature sets, the nuclear segmentations from Section 2.1 were used to calculate the cell boundaries and centroids. In total, we investigated the performance of 4 feature cohorts: (1) 100 features describing cell morphology, (2) 51 features describing cell architectures, (3) 39 features describing cell orientation entropy (CORe), and (4) the combined feature set spanning cohorts (1-3).

3.3 Random Forest Classifier

In this study, we demonstrate the efficacy of including CORe features for improving classification accuracy and area under the receiver operating characteristic curve (AUC) in predicting BCR in CaP patients from prostate TMAs. Randomized 3-fold cross validation was performed on the top 10 most informative features selected via Student *t*-test for each of 4 feature cohorts defined in Section 3.2. Classification was performed using a random forest classifier.

Table 2. Summary of 151 nuclear morphologic features

Cell Morphology	#	Description
	100	Area Ratio, Distance Ratio, Standard Deviation of Distance, Variance of Distance, Distance Ratio, Perimeter Ratio, Smoothness, Invariant Moment 1-7, Fractal Dimension, Fourier Descriptor 1-10 (Mean, Std. Dev, Median, Min / Max of each)
Cell Architecture		Description
Voronoi Diagram	12	Polygon area, perimeter, chord length: mean, std. dev., min/max ratio, disorder
Delaunay Triangulation	8	Triangle side length, area: mean, std. dev., min/max ratio, disorder
Minimum Spanning Tree	4	Edge length: mean, std. dev., min/max ratio, disorder
Nearest Neighbors	27	Density of nuclei, distance to nearest nuclei

4 Results and Discussion

Figure 1 reveals the ability of the CORe features to capture the differences in angular disorder across localized cell networks and illustrates the differences between the BCR and NR cases in terms of the CORe features.

In Table 3, we can summarize the performance of feature descriptors describing cell architecture and cell morphology which appear to have a maximum BCR prediction accuracy of 79.9%. However, by inclusion of novel cell orientation entropy (CORe) features, the overall classifier accuracy improves to 82.7%. Similar improvements are also observed in terms classification AUC. This reflects the utility of CORe features as a valuable prognostic measurement for predicting BCR in conjunction with previously described nuclear morphologic features.

Classifier improvement following inclusion of CORe features suggests that many of the new CORe features are non-correlated with previously defined cell architectural and morphological feature sets. This distinction is illustrated in Figure 1, where we observe the differences between CORe features compared with those obtained from Voronoi and Delaunay graphs. These graphs span across stromal and epithelial regions, while CORe features are limited to sub-graphs in localized regions. It is also important to note that the combination of CORe and nuclear morphologic features clearly and significantly outperform the clinical standard of pathologist grade, which classified all cases as GS 7.

Table 3. 100 runs of 3-fold Random Forest Classification

	Architecture	Morphology	CORe	Arch + Morph + CORe
Accuracy	71.2 ± 4.2%	79.9 ± 3.7%	74.6 ± 4.1%	82.7 ± 3.1%
AUC	0.641 ± 0.054	0.773 ± 0.042	0.688 ± 0.063	0.809 ± 0.037

5 Concluding Remarks

In this work, we presented a new feature descriptor, cell orientation entropy (COrE), for quantitative measurement of local disorder in nuclear orientations in digital pathology images. We demonstrated high accuracy and improvement in predicting BCR in 39 CaP TMAs via the use of COrE features. While COrE features did not outperform other quantitative histomorphometric measurements such as nuclear shape and architecture significantly, the combination of nuclear shape, architectural and COrE features boosted classifier accuracy in identifying patients at risk for BCR following radical prostatectomy. More significantly, the combination of COrE and other image based features significantly outperformed pathologist derived GS, which is 50% for GS 7, and is further known to have at best moderate inter-observer agreement ($\kappa = 0.47\text{--}0.7$) [3]. In future work, we aim to evaluate the applicability of COrE features in other disease sites such as breast cancer.

References

1. Trock, B., et al.: Prostate cancer-specific survival following salvage radiotherapy vs observation in men with biochemical recurrence after radical prostatectomy. *JAMA* 299(23), 2760–2769 (2008)
2. Han, M., et al.: Biochemical (prostate specific antigen) recurrence probability following radical prostatectomy for clinically localized prostate cancer. *J. Urol.* 169(2), 517–523 (2003)
3. Allsbrook Jr., W., et al.: Interobserver reproducibility of gleason grading of prostatic carcinoma: general pathologist. *Hum. Pathol.* 32(1), 81–88 (2001)
4. Christens-Barry, W., Partin, A.: Quantitative grading of tissue and nuclei in prostate cancer for prognosis prediction. *Johns Hopkins Apl Technical Digest* 18, 226–233 (1997)
5. Tabesh, A., et al.: Multifeature prostate cancer diagnosis and gleason grading of histological images. *IEEE Transactions on Medical Imaging* 26(10), 1366–1378 (2007)
6. Jafari-Khouzani, K., Soltanian-Zadeh, H.: Multiwavelet grading of pathological images of prostate. *IEEE Trans. on Biomedical Engineering* 50(6), 697–704 (2003)
7. Bilgin, C., et al.: Cell-graph mining for breast tissue modeling and classification. *IEEE Eng. in Med. and Biol. Soc., EMBS*, 5311–5314 (2007)
8. Veltri, R.W., et al.: Nuclear roundness variance predicts prostate cancer progression, metastasis, and death: A prospective evaluation with up to 25 years of follow-up after radical prostatectomy. *The Prostate* 70(12), 1333–1339 (2010)
9. Epstein, J.I.: An update of the gleason grading system. *Journal Of Urology (the)* 183(2), 433 (2010)
10. Haralick, R., et al.: Textural features for image classification. *IEEE Trans. on Systems, Man and Cybernetics* (6), 610–621 (1973)
11. Ali, S., Madabhushi, A.: An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE Transactions on Medical Imaging* 31(7), 1448–1460 (2012)
12. Doyle, S., et al.: Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer. *BMC Bioinformatics* 13(1), 282 (2012)