2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Evaluation of Feature Descriptors for Cancerous Tissue Recognition

Panagiotis Stanitsas[1], Anoop Cherian[2], Xinyan Li[1], Alexander Truskinovsky[3], Vassilios Morellas[1] and Nikolaos Papanikolopoulos[1]

[1]Department of Computer Science and Engineering, University of Minnesota
[2]Australian Center for Robotic Vision, Australian National University
[3]Department of Pathology & Laboratory Medicine, Roswell Park Cancer Institute

***Abstract*—Computer-Aided Diagnosis (CAD) has witnessed a rapid growth over the past decade, providing a variety of automated tools for the analysis of medical images. In surgical pathology, such tools enhance the diagnosing capabilities of pathologists by allowing them to review and diagnose a larger number of cases daily. Geared towards developing such tools, the main goal of this paper is to identify useful computer vision based feature descriptors for recognizing cancerous tissues in histopathologic images. To this end, we use images of Hematoxylin & Eosin-stained microscopic sections of breast and prostate carcinomas, and myometrial leiomyosarcomas, and provide an exhaustive evaluation of several state of the art feature representations for this task. Among the various image descriptors that we chose to compare, including representations based on convolutional neural networks, Fisher vectors, and sparse codes, we found that working with covariance based descriptors shows superior performance on all three types of cancer considered. While covariance descriptors are known to be effective for texture recognition, it is the first time that they are demonstrated to be useful for the proposed task and evaluated against deep learning models. Capitalizing on Region Covariance Descriptors (RCDs), we derive a powerful image descriptor for cancerous tissue recognition termed, Covariance Kernel Descriptor (CKD), which consistently outperformed all the considered image representations. Our experiments show that using CKD lead to 92.83%, 91.51%, and 98.10% classification accuracy for the recognition of breast carcinomas, prostate carcinomas, and myometrial leiomyosarcomas, respectively.**

I. Introduction

An essential step towards the successful treatment of cancer is its early and accurate diagnosis. This requires close examination of tissue slides from suspected regions under a microscope – a task which is often very time consuming, thus limiting the number of cancer cases that experts can handle daily. Given that hospitals and clinics are facing a continuously increasing number of such cases, while the number of expert pathologists for the task is limited, it is clear that automated tools, with the ability to confidently identify prospective cancerous regions, can assist the pathologists and immensely speed up the diagnosis.

The goal of this paper is the development of a CAD scheme for expediting the analysis of Hematoxylin & Eosin (H&E)-stained tissue samples. H&E staining is a commonly used technique in pathology where Hematoxylin will stain the nuclei in blue or dark purple color, while Eosin imparts a pink or lighter purple color to the cytoplasm, as depicted in Figure 1

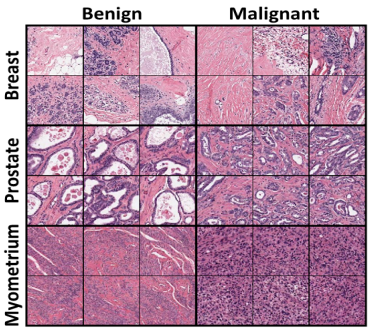


Fig. 1: Hematoxylin & Eosin-stained samples for three types of tissue; Breast (1st and 2nd row), Prostate (3rd and 4th rows) and Myometrium (5th and 6th row).

for the considered types of tissue. An automated identification of the regions that are highly likely to be cancerous, can assist experts in finding them among the surrounding benign tissues efficiently, resulting in faster diagnosis.

To achieve this goal, in this paper, we study discriminative image representations that can confidently classify a benign image patch from a cancerous one. In pursuit of such a representation, we evaluate several state-of-the-art feature descriptors that have demonstrated significant promise for mainstream computer vision applications, including object recognition, texture recognition, and shape discrimination (Section IV). We compare classical feature descriptors such as Histograms of Oriented Gradients (HOG) and Gabor wavelets, as well as more recent representations based on Convolutional Neural Networks (CNN), Fisher Vectors (FVs), sparse codes and Region Covariance Descriptors (RCDs). Of these, RCD, which fuse raw image features (such as image intensity and gradients) into a compact positive definite matrix, is perhaps the simplest to generate (Section IV-B). We also derive an extension of the RCD, dubbed Covariance-Kernel Descriptor (CKD), by combining it with a positive definite kernel matrix generated from color histograms. Moreover, our evaluation shows that RCD and CKD, when combined with a suitable non-linear geometry, can in fact offer superior classification performance for the proposed task against other descriptors. To the best of our knowledge, the application of RCD and CKD for cancer tissue recognition has not been investigated in the past.

A major advantage of the proposed approach is the fact that

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

segmentation of the nuclei is not required as a pre-processing step since global image descriptors are used. This allows our scheme to operate without being constrained by grading systems (e.g., Gleason grading system for prostate cancer), making it easily extensible to other types of cancer by a proper training procedure.

To evaluate our algorithms, we construct datasets for three types of cancer, namely (i) breast, (ii) prostate, and (iii) myometrium. To this end, microscopic images from H&E-stained sections from malignant and benign regions are used for each of these tissue types. Our data collection process is described in Section III. Extensive comparisons of the various feature representations using different evaluation metrics are presented in Section V. To set the stage for our discussions, in the next section, we briefly review some of the prior computer vision based approaches to cancerous tissue recognition.

II. Related Work

Several techniques have been presented over the past decade for the accurate detection of cancerous segments in various types of medical images. Classification of cancerous regions on histopathological images can be performed at the level of the tissue architecture, as well as at the nuclear level. In general, the intensity of each pixel and its spatial dependence is used as an image descriptor. These features could be further categorized ([8], [15]) based on: 1) intensity (i.e., density, hue, mean, median, and variance), 2) morphology (i.e., area, center of mass, concavity, minor axis, and major axis), and 3) texture (i.e. co-occurrence features, fractal dimension, run-length feature, and Gabor wavelets).

Run length is defined as the number of consecutive pixels with the same intensity value in a given direction. Features are extracted from the gray-level run-length matrix, which is then used to count the total occurrences. Sun et al. [19] propose such a system for prostate cancer detection using a run-length matrix. Feature co-occurrences is another commonly used strategy for generating descriptors for cancer detection, first introduced in [9]. Systems have been developed to classify liver cancer [10], as well as prostate cancer (e.g. [7], [16]), relying on co-occurrence features.

Other frequently used features are based on signal processing, such as filtering and transformations to the frequency domain. For example, Sobel filters, in the x, y, and two diagonal axes, are used in [7]. In Nguyen et al. [16], the Kirsch filter, as well as gradients in the x and y directions are suggested. Doyle et al. [7] suggests Gabor wavelets for discriminating cancer tissues. Cruz-Roa et al. [3] presented a methodology capitalizing on Deep Learning, while comparisons were also established with Bag of Visual Words representations and Haar features. Liu et al. [13] had also performed a comparative study on both morphological and texture features in order to explore the optimal features for nuclei classification including Daubechies and Gabor Wavelets.

Once the suitable features are selected, standard machine learning based classification schemes can be used for cancer diagnosis. For example, Alexandratou et al. [1] conducted a large scale comparison between sixteen classification algorithms for prostate cancer diagnosis. Furthermore, an ensemble of two SVMs was presented in Nguyen et al. [16] for detecting prostate cancer using cytological and textural features.

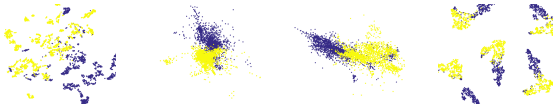


Fig. 2: A low dimensional embedding using tSNE for the myometrium dataset using (a) Normalized Color Histograms (b) Covariance Descriptors (c) Covariance Kernel Descriptors (d) CNN features.

Given that general-purpose computer vision feature representations have demonstrated significant promise for very challenging real-world vision tasks, we believe it is essential to understand their performance on medical images, which offer a more controlled and high-impact setting. To this effect, our goal is to evaluate the best available image representations in a systematic way on a diverse collection of data in the context of medical imaging and draw valuable conclusions about the information content that they manage to capture.

III. Data Collection

In this study, data from three types of cancer are used to assess the performance of the proposed algorithms, namely carcinomas of the prostate, the breast, and the myometrium. The tissue samples collected are (H&E)-stained, followed by high-resolution (10K $\times$ 9K pixels) scans of tissue sections taken at x50 magnification on a digital slide scanner. A medical expert (surgical pathologist) was responsible for providing annotations corresponding to the malignant and benign image regions. The annotated regions are then divided into smaller disjoint patches of 150×150 pixels.

Next, binary class labels are assigned to each of the image patches. That is, those patches for which more than 80% of the pixels correspond to carcinomas, are treated as the positive class, while patches in the negative class are devoid of any cancerous regions. For the case of prostate cancer, 31 images of carcinomas and 8 images from benign regions are annotated, taken from 10 patients. We then construct a balanced dataset for training and testing purposes, containing 3500 image patches with 1750 patches depicting cancerous regions, while the other 1750 corresponding to benign regions. For the case of carcinomas of the breast, we used 21 annotated images of carcinomas and 19 images of benign tissue, taken from 21 patients. Similarly we construct a dataset of 3500 randomly selected image patches of which, 1750 depicted cancerous cases while the other half corresponded to benign cases. Finally, 39 myometrial leiomyomas were combined with 41 images of leiomyosarcomas to construct the third dataset from 39 patients. We randomly selected 1539 cancerous image patches and combined them with 1782 benign patches to total a dataset of 3321 samples.

IV. Image Representations

Towards an accurate classification between benign and malignant tissues, several types of feature representations

are evaluated in this section. We first consider the naïve representation using the raw pixel intensities of gray-scale image patches. In that way, for an $n \times n$ image patch, a vectorial representation of size n^2 is derived by concatenating the columns of the patch. It appears that such a representation fails to capture invariances (such as to pixel color and spatial locations) that are useful for classification between benign and malignant tissue types. This is substantiated by training a classifier on such vectorized images (as reported in Table I). Having learned from this shortcoming, in the sequel, we investigate more powerful feature representations.

A. Normalized Color Histograms

In an effort to further explore the information uncovered by the H&E staining, Normalized Color Histograms (NCH) were computed. We computed color histograms consisting of 256 bins each for the R, G, and B color channels; this histogram is normalized to sum to one and concatenated to form a 768-dimensional feature descriptor for the respective patch. To intuitively understand this representation, we plot in Figure 2(a), a low-dimensional embedding of these features using the t-Distributed Stochastic Neighbor Embedding (t-SNE) [6] method, which depicts a coherent cluster formation as also supported by our experimental evaluation (Table I).

B. Region Covariance Descriptors

RCDs have been proposed for many applications in computer vision. In contrast to the typical high-dimensional feature descriptors that often assume a flat Euclidean geometry, RCDs are generally low-dimensional and are assumed to belong to a highly non-linear geometry. In their basic form, RCDs are generated as described in Equation (1), where $\mathbf{f_i} \in \mathbb{R}^d$, are d-dimensional features extracted from each pixel $i \in \{1, 2, \cdots, N\}$ of an image patch, and μ is the mean feature given by $\mu = \frac{1}{N}\sum_{i=1}^{N} \mathbf{f_i}$.

$$\mathbf{C} = \frac{1}{(N-1)} \sum_{i=1}^{N} (\mathbf{f_i} - \mu)(\mathbf{f_i} - \mu)^T. \quad (1)$$

RCDs are covariance matrices computed over a set of features extracted from every pixel in the image patch. In this paper we consider a 5-dimensional RCD consisting of the normalized intensities of the three channels $\mathbf{R}$, $\mathbf{G}$, and $\mathbf{B}$ of a color image combined with first-order gradient information along the x and y axis, as denoted by $\mathbf{G_i^x}$ and $\mathbf{G_i^Y}$ respectively. That is, our f_i has the following form:

$$\mathbf{f_i} = [\mathbf{R_i} \;\; \mathbf{G_i} \;\; \mathbf{B_i} \;\; \mathbf{G_i^x} \;\; \mathbf{G_i^y}]^T. \quad (2)$$

Covariance matrices are symmetric positive definite (SPD) matrices. Given that SPD matrices form an open subspace of the Euclidean space, it is natural to assume a Euclidean geometry to these matrices. However, it is often found that [17] assuming a non-linear geometry is often beneficial practically. That is, instead of using a Euclidean distance to measure the similarity between two SPD matrices, a non-linear measure is used which governs the geometry of the space of these matrices. Two commonly used such measures are (i) the Log-Euclidean Riemannian metric, and the recently introduced (ii) Jensen-Bregman Logdet Divergence. Of these two, (i) defines a Riemannian geometry to the space of SPD matrices, while (ii) defines an information geometry based similarity measure.

First, the Log-Euclidean Riemannian Metric (LERM) is described in Equation 3 for a pair of covariance matrices $\mathbf{C^{(i)}}$ and $\mathbf{C^{(j)}}$. In Riemannian geometry, the set of symmetric matrices form a tangent space for the Riemannian manifold of SPD matrices, and the space of symmetric matrices is isomorphic to the Euclidean space. Thus, taking the matrix logarithm, as in (3), embeds the SPD matrices into a flat tangent space of symmetric matrices on which the usual Euclidean distance can be used for similarity computations.

$$D_{LERM}(\mathbf{C^{(i)}}, \mathbf{C^{(j)}}) := \left\| \mathrm{Log}(\mathbf{C^{(i)}}) - \mathrm{Log}(\mathbf{C^{(j)}}) \right\|_{\mathbf{F}}, \quad (3)$$

where $\mathrm{Log}(\cdot)$ is the matrix logarithm and $\|\cdot\|_{\mathbf{F}}$ is the Frobenius norm.

Second, the Jensen-Bregman LogDet Divergence (JBLD), first proposed by Cherian et al. [2], is also considered for similarity computations, as presented in Equation (4). In contrast to LERM, JBLD retains the rich non-linear geometry of the space of SPD matrices, and at the same time is computationally cheaper as the matrix logarithms are replaced by matrix determinants which can be computed efficiently via Cholesky factorization. Computing a 2-dimensional embedding for the myometrium dataset, can visually support the meaningful cluster formation when capitalizing on RCDs, as depicted in Figure 2(b).

$$D_{JBLD}(\mathbf{C^{(i)}}, \mathbf{C^{(j)}}) := \left[log \left| \frac{\mathbf{C^{(i)}} + \mathbf{C^{(j)}}}{2} \right| - \frac{1}{2} log \left| \mathbf{C^{(i)}} \mathbf{C^{(j)}} \right| \right]^{1/2}, \quad (4)$$

where $|A|$ is the determinant of SPD matrix A.

C. Covariance-Kernel Descriptors

Capitalizing on the information captured by the RCDs and the NCHs, we combine the two representations towards deriving a stronger descriptor. Recall that the RCDs compute the feature correlations between each pixel in the patch against other pixels; thus capturing texture and shape in the patch implicitly. However, RCDs make an implicit dependency between the attributes of a pixel and the pixel location in the patch. While this dependency is important for cancerous tissue recognition, sometimes spatial invariance of the color histograms is more important as suggested by the NCH descriptor above. Thus, both RCDs and NCHs capture complementary cues for recognition, and thus we expect their combination to provide a synergy to the overall accuracy.

Motivated by the above intuition, we propose a novel fusion of RCDs and NCHs to generate a Covariance-Kernel Descriptor (CKD) as follows. We generate a compact block diagonal symmetric positive definite matrix descriptor that contains in its first block the RCD denoted by $\mathbf{C}$ as computed in Equation (1), while the second block captures the correlations between the histograms computed on the three

color channels of the image patch (as in the NCH). However, rather than concatenating the three histograms, as presented in Section IV-A, we combine them to formulate a matrix $\mathbf{H} \in \mathbb{R}^{3\times b}$, where each row corresponds to the b-bin histogram on a channel. The resulting CKD matrix is as follows:

$$\mathbf{CKD} = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}\mathbf{H}^T \end{bmatrix}, \quad (5)$$

where $\mathbf{0}$ is a 3×3 matrix of all zeros.

Given that the 3x3 histogram correlation matrix HH^T is positive definite (and thus a valid Mercer kernel), we could further improve its representational power by computing the correlations via a kernel function. That is, suppose $h_c \in \mathbb{R}^b$ denotes a histogram vector (where $c \in \{R, G, B\}$), then we could replace the Gram matrix HH^T in (5) by a kernel matrix K defined by $K(h_{c1}, h_{c2}) = \phi(h_{c1})^T\phi(h_{c2})$ for $c1, c2 \in \{R, G, B\}$ and a kernel function ϕ. However, we found that for our task, the linear kernel performed the best and thus we use this setup in our evaluation. Note that the resulting fused descriptor is still an SPD matrix and thus we use the similarity measures defined for the RCD for CKD as well. A 2-dimensional embedding for the dataset of the myometrium depicts the cluster separability when capitalizing on CKDs, as depicted in Figure 2(c).

D. Bag of Visual Words

Bag Of visual Words (BOW) representation relies on the creation of a codebook which compactly characterizes a collection of local point features [4]. In this paper, we used BOW on Scale Invariant Feature Transform (SIFT) descriptors. For each type of tissue, we randomly select 1000 image patches and compute a collection of 128 dimesional SIFT descriptors for each patch. For generating the codebook for each type of cancer, we cluster the extracted SIFT descriptors using K-Means in 300 clusters. The resulting centroids after clustering are used as the codewords. To encode a patch in terms of the generated codebook, we first extract its SIFT descriptors from the patch, followed by generating a 300 dimensional histogram depicting the frequency by which these descriptors are assigned to the most similar codewords. This histogram is normalized to sum to one, thus depicting a discrete probability distribution, and is used as an encoding of the image patch.

E. Sparse Codes

Sparse coding has revolutionized the domain of machine learning and computer vision by enabling the construction of informative representations of data as linear (in most cases) combinations of a few representative "learned" atoms (e.g. [14]). For the purposes of our study, we start by learning a such matrix of atoms, dubbed a dictionary matrix, for each type of tissue using the SPAMS library[1]. In this study, the size of the patches (150×150) is prohibitive to allow learning a dictionary. To circumvent this issue, we divided each patch into 100 smaller patches of size 15×15. Working with this 225-dimensional vectorized representations for the smaller patches, we learn a dictionary of 1500 atoms. Note that we normalized the data to have zero mean and unit norm before training the dictionary. Given a test patch of size 150×150, we repeat the process and generate 1500-dimensional sparse codes for each 15×15 patch by solving a Lasso objective. This precedes aggregating the sparse codes via average pooling, thus generating 1500-dimensional descriptors for the full image patch.

F. Gabor Features

Gabor based features have been previously shown to be useful for cancerous tissue recognition (e.g., [13]). In this study, we generated a bank of Gabor filters at different orientations and scales. Particularly, we present results based on a Gabor space constructed by convolutions in 4 orientations (0°, 45° and 90°, 135°) and 4 scales with a kernel size of 7×7 pixels. These parameters were selected via cross-validation on a subset of our dataset. After gray-scale images are convolved with the selected filters, they are downsampled by a factor of 4 and vectorized. Finally, the resulting vectors for all filters are concatenated together to form a 23,104-dimensional descriptor.

G. Histogram of Oriented Gradients (HOG)

HOG descriptors [5] are classic computer vision descriptors that can capture shapes in images and have demonstrated to be immensely useful for object recognition and person detection tasks. A HOG descriptor is generated by dividing an image patch into a set of non-overlapping cells, followed by computing a histogram of intensity gradients in each cell. In our case, through a trial and error process we selected to work with a cell size of 6×6 pixels, while 31 bins are used to produce the histogram for each cell. The VL-FEAT toolbox [2] was utilized to compute the HOG descriptors for our experiments based on the aforementioned characteristics, producing a 19,375-dimensional descriptor.

H. Fisher Vectors

Fisher vectors (FVs) provide a significant enhancement over the BOW model in a probabilistic/information geometric [18]. Instead of using a hard clustering algorithm (such as K-Means) on the SIFT descriptors, FV uses probabilistic clustering using Gaussian Mixture Models (GMM). Further, the main insight in the development of these descriptors is the observation that the gradient of the log-likeligood of the GMM with respect to the parameters of the component Gaussian distributions provides the direction in which the model parameters need to be adjusted to better approximate the data. This gradient is also related to the Fisher information matrix when the space of GMMs is regarded as a Riemannian manifold (and hence the name). In our experiments, we used 300 Gaussians to represent our feature descriptors which resulted in a 76800-dimensional representation. Once again the VL-FEAT[2] toolbox was used for our computations.

[1] http://spams-devel.gforge.inria.fr/

[2] http://www.vlfeat.org/

I. Deep Learning

The main pursuit of deep Convolutional Neural Networks (CNNs) is learning optimal transformations of the data that enhance the separability between classes. For a concise outline of the domain on Deep Learning and CNNs we refer the reader to [12]. However, CNNs consist of millions of parameters and thus demand large corpus of data to train them effectively, which can be daunting for tasks such as ours. Since we have data limited to a few thousand samples, we fine tune a pre-trained CNN model. In that way, we allow the fully connected layers of the network to continue learning while the convolutional layers are restricted from learning at the same pace by significantly lessening their learning rates.

Inspired by the high accuracy demonstrated on the Imagenet object classification benchmarks, for this study, we used the popular Alexnet topology [11] within the Caffe framework[3]. In addition to the demonstrated accuracy, this topology is also significantly less demanding on GPU memory, thus avoiding the need for sophisticated hardware. A 2-dimensional embedding for the myometrium dataset, visually supports the informativeness of representations generated by the CNN, as depicted in Figure 2(d).

V. Experiments

To assess the discriminatory power of the selected representations, we conduct a series of experiments within a supervised classification framework. In particular, we present comparisons using 5-Nearest Neighbors (5-NN) classifiers, Support Vector Machines (SVMs), as well as a linear classifier at the last layer of the deployed CNN. For all the learned models, we evaluate the classification performance using two different metrics, namely (i) classification accuracy (ACC), and (ii) the Area Under the Curve (AUC) computed from Receiver Operating Characteristic (ROC) curves. To produce more generalizable conclusions we used a 10-fold cross-validation for all our experiments.

For our SVM based experiments, we used the popular libSVM library[4]. For RCDs and CKDs, we use Radial Basis Function (RBF) Mercer kernels based on the LERM and the JBLD measures. For the rest of the tested descriptors, a collection of different kernels and parameter configurations were tested. In particular, the tested kernels were linear, polynomial, RBF and Sigmoid. For almost all feature representations, linear kernels achieved the highest performance and were used to report our results. The only exception is the kernel utilized for the Gabor features which is a polynomial kernel of 3rd degree.

Finally, for the CNN we slightly alter the topology of the network to reflect the number of classes of the problem in hand, which is 2. In its original implementation, the number of classes was 1000. Since training a network from scratch is prohibitive given the limited amount of data, we capitalize on a pre-trained network and finetune it. This was achieved by significantly decreasing the learning rates of the convolutional layers of our network and allowing mostly the inner product layers to continue learning based on our dataset. We also experimented with different learning rates with no significant impact on the performance. We initialize the weights of our network with weights learned on the 1M image database of the ILSVRC challenge and we perform additional 5K iterations, which were shown to be sufficient for the problem in hand.

[3]http://caffe.berkeleyvision.org/

[4]https://www.csie.ntu.edu.tw/ cjlin/libsvm/

TABLE I: Experimental Results.

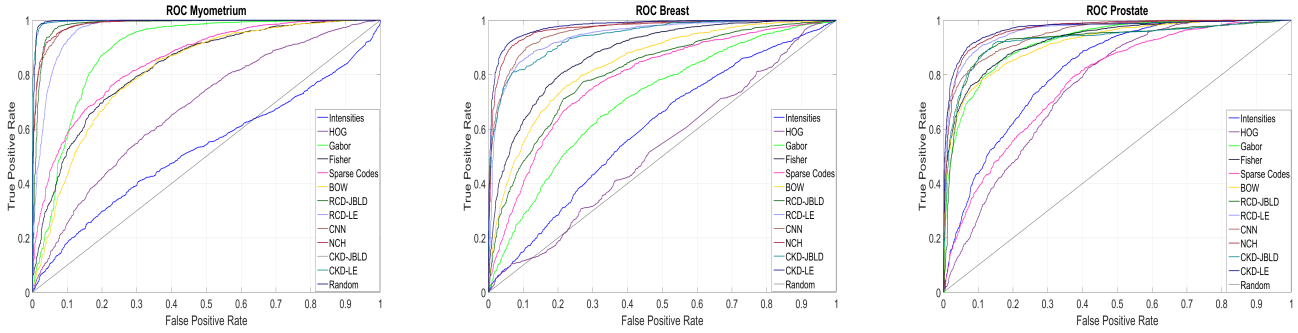
	Myometrium		Breast		Prostate	
Features | Classifier	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**
Intensities | 5-NN	46.33%	–	49.69%	–	69.54%	–
Intensities | SVM	50.51%	0.53	57.91%	0.60	73.71%	0.82
HOG | 5-NN	55.72%	–	60.14%	–	48.23%	–
HOG | SVM	62.89%	0.68	51.86%	0.53	69.51%	0.76
Gabor | 5-NN	46.60%	–	52.09%	–	48.66%	–
Gabor | SVM	84.37%	0.89	65.60%	0.71	83.54%	0.92
Fisher | 5-NN	59.31%	–	63.49%	–	75.57%	–
Fisher | SVM	74.91%	0.83	79.66%	0.88	84.37%	0.93
Sparse Codes | 5-NN	55.10%	–	67.51%	–	51.37%	–
Sparse Codes | SVM	76.54%	0.85	72.31%	0.78	69.66%	0.77
BOW | 5-NN	56.63%	–	66.03%	–	67.06%	–
BOW | SVM	74.85%	0.81	76.46%	0.84	83.09%	0.92
RCD-JBLD | 5-NN	92.53%	–	67.06%	–	79.09%	–
RCD-JBLD | SVM	95.24%	0.98	74.26%	0.81	87.29%	0.92
RCD-LE | 5-NN	91.81%	–	67.09%	–	79.66%	–
RCD-LE | SVM	91.93%	0.97	87.66%	0.94	89.77%	0.96
CNN(AlexNet)	93.77%	0.99	89.23%	0.96	86.91%	0.95
NCH | 5-NN	95.03%	–	84.60%	–	82.00%	–
NCH | SVM	93.91%	0.99	91.63%	0.97	90.26%	0.96
CKD-JBLD | 5-NN	95.30%	–	79.31%	–	80.06%	–
CKD-JBLD | SVM	97.86%	1	85.51%	0.94	86.63%	0.93
CKD-LE | 5-NN	94.88%	–	79.51%	–	80.66%	–
CKD-LE | SVM	**98.10%**	**1**	**92.83%**	**0.98**	**91.51%**	**0.97**

VI. Discussion of Results and Future Work

In order to facilitate this discussion, we aggregate the results in Table I for all the described feature representations in terms of ACC and AUC, as computed for the extracted ROC curves. Figure 3 presents the resulting ROC curves for all the conducted classification experiments.

Based on our results we can infer that the tested descriptors that use color information perform better against those that are extracted based only on gray-scale intensities. This latter category of descriptors includes, gray-scale intensities, HOG, FVs, Gabor wavelets, sparse codes and BOW. Among those, FVs appear to achieve the highest accuracy as well as AUC, reaching accuracy of 74.91%, 79.66% and 84.37% for the myometrium, breast, and prostate dataset, respectively. This, though, comes with a computational overhead, derived from the large dimensionality of the descriptor. The NCH was the only feature representation that was built solely on color information. Nevertheless, this was shown to be sufficient to outperform all the aforementioned edge-based descriptors and was only exceeded by descriptors using both edge and color information. NCH achieved accuracy values reaching 93.91%, 91.63% and 90.26% for the myometrium, breast, and prostate dataset, respectively, accompanied by very high AUC. The achieved performances, combined with the low dimensionality and ease of computation makes this descriptor a very attractive solution for cancer recognition tasks on H&E stained images.

In the final step of this experimentation, descriptors balancing both color and gradient information were considered. In particular, RCDs and CNN reported accuracies that were on par with the performance of the NCHs. RCDs exceeded



(a) (b) (c)

Fig. 3: ROC curves for (a) Myometrium, (b) Breast and (c) Prostate (Best if viewed in color).

the performance on myometrium by 1.33% but in the case of breast and prostate cancer they achieved a lower accuracy of 87.66% and 89.77% respectively. For the myometrium and prostate datasets, CNN representations achieved a lower accuracy (93.77% and 86.91% respectively) both compared to RCDs and NCHs. For the breast carcinoma, although CNNs exceeded the performance of RCDs (89.23%), they did not perform better than NCHs.

Finally, CKD, a descriptor introduced in this work, is seen to outperform all the considered descriptors, reaching ACC of 98.1%, 92.83% and 91.51% for the myometrium, breast and prostate dataset, respectively. The enhanced pixel intensity invariance infused by the color histogram Gram matrix, along with the gradient information and spatial correlation of pixel values integrated by the RCDs allowed this descriptor to reach an AUC value of almost 1 for the myometrium dataset.

Collectively, we have presented a methodical feature evaluation for a large collection of general-purpose computer vision feature representations on three types of cancer. Furthermore, we introduce two descriptors, RCDs and CKDs in the context of cancerous tissue recognition. CKDs were able to outperform all the tested representations including the deployed CNN scheme. The presented methodology will be expanded to additional types of tissue, including the colon, pancreas, lung, and others as more annotated data becomes available.

VII. Acknowledgments

We would like to thank the anonymous reviewers for their thoughtful comments. This material is based upon work supported by the National Science Foundation through grants #IIP-0934327, #CNS-1039741, #SMA-1028076, #CNS-1338042, #IIP-1439728, #OISE-1551059, and #CNS-1514626. Dr. Cherian is funded by the Australian Research Council Centre of Excellence for Robotic Vision (project number CE140100016).

References

[1] E. Alexandratou, V. Atlamazoglou, T. Thireou, G. Agrogiannis, D. Togas, N. Kavantzas, E. Patsouris, and D. Yova. Evaluation of machine learning techniques for prostate cancer diagnosis and gleason grading. *International Journal of Intelligent Bioinformatics Systems Biology*, 1(3):297–315, 2010.

[2] A. Cherian, S. Sra, A. Banerjee, and N. Papanikolopoulos. Jensen-bregman logdet divergence with application to efficient similarity search for covariance matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2161–2174, 2013.

[3] A. Cruz-Roa, J. Ovalle, A. Madabhushi, and F. Osorio. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. In *Medical Image Computing and Computer-Assisted Intervention*, pages 403–410. Springer, 2013.

[4] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, European Conference on Computer Vision*, volume 1, pages 1–2, 2004.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, 2005.

[6] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.

[7] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi. A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. *IEEE Transactions on Biomedical Engineering*, 59(5):1205–1218, 2012.

[8] L. Roux H. Irshad, A. Veillard and D. Racoceanu. Methods for nuclei detection, segmentation, and classification in digital histopathology: A review – current status and future potential. *IEEE Reviews in Biomedical Engineering*, 7:97–114, 2014.

[9] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, 1973.

[10] Po-Whei Huang and Yan-Hao Lai. Effective segmentation and classification for {HCC} biopsy images. *Pattern Recognition*, 43(4):1550 – 1563, 2010.

[11] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[12] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[13] S. Liu, P. A. Mundra, and J. C. Rajapakse. Features for cells and nuclei classification. In *IEEE International Conference on Engineering in Medicine and Biology Society*, pages 6601–6604, 2011.

[14] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In *ACM International Conference on Machine Learning*, pages 689–696, 2009.

[15] C. Mosquera-Lopez, S. Agaian, A. Velez-Hoyos, and I. Thompson. Computer-aided prostate cancer diagnosis from digitized histopathology: A review on texture-based systems. *IEEE Reviews in Biomedical Engineering*, 8:98–113, 2015.

[16] K. Nguyen, A. Jain, and B. Sabata. Prostate cancer detection: Fusion of cytological and textural features. *Journal of Pathology Informatics*, 2(2):3, 2011.

[17] X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

[18] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek. Image classification with the fisher vector: Theory and practice. *International Journal of Computer Vision*, 105(3):222–245, 2013.

[19] X. Sun, S. Chuang, J. Li, and F. McKenzie. Automatic diagnosis for prostate cancer using run-length matrix method. In *SPIE Medical Imaging*, pages 72603H–72603H, 2009.