

Stacked Predictive Sparse Coding for Classification of Distinct Regions in Tumor Histopathology *

Hang Chang^{1*†} Yin Zhou^{1†} Paul Spellman² Bahram Parvin^{1*}

¹Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, U.S.A

{hchang, yinzhou, b_parvin}@lbl.gov, * Co-Corresponding authors, † Co-First authors

² Center for Spatial Systems Biomedicine, Oregon Health Sciences University, Portland, Oregon, U.S.A

spellmap@ohsu.edu

Abstract

Image-based classification of histology sections, in terms of distinct components (e.g., tumor, stroma, normal), provides a series of indices for tumor composition. Furthermore, aggregation of these indices, from each whole slide image (WSI) in a large cohort, can provide predictive models of the clinical outcome. However, performance of the existing techniques is hindered as a result of large technical variations and biological heterogeneities that are always present in a large cohort. We propose a system that automatically learns a series of basis functions for representing the underlying spatial distribution using stacked predictive sparse decomposition (PSD). The learned representation is then fed into the spatial pyramid matching framework (SPM) with a linear SVM classifier. The system has been evaluated for classification of (a) distinct histological components for two cohorts of tumor types, and (b) colony organization of normal and malignant cell lines in 3D cell culture models. Throughput has been increased through the utility of graphical processing unit (GPU), and evaluation indicates a superior performance results, compared with previous research.

1. Introduction

Tumor histology provides a detailed insight into cellular morphology, organization, and heterogeneity. For example, histology sections can be used to identify mitotic cells, cellular aneuploidy, and autoimmune responses. More importantly, if tumor morphology and architecture can be quantified in a large cohort, it will provide the basis for predictive models in a similar way that genomic techniques have identified predictive molecular subtypes. Genome wide analysis

techniques (e.g., microarray analysis) have the advantages of standardized tools for data analysis and pathway enrichment, which enables hypothesis generation for the underlying mechanism. On the other hand, histological signatures are hard to compute because phenotypic signatures are not standardized and advanced methods for image analysis remain at a deficit. Image analysis is further complicated by technical variations as a result of sample preparation (e.g., fixation, staining) and biological heterogeneity, where the latter originates within a whole slide image (WSI) and between WSIs as no two patients have the same aberrant signature.

Although manual analysis may incur inter- and intra-pathologist variations [12] and some researchers have focused on reducing such variations, the value of the quantitative histological image analysis rests on its capability in capturing detailed morphometric features and organization. Such rich descriptions can then be linked with genomic information and clinical outcomes for improved diagnosis and therapy.

One of the main technical barriers for processing a large collection of histological data is that the color composition is subject to technical variations (e.g., fixation, staining) and biological heterogeneities (e.g., cell type, cell state) across histological tissue sections, especially when these tissue sections are processed and scanned at different laboratories. Here, a histological tissue section refers to an image of a thin slice of tissue applied to a microscopic slide and scanned from a light microscope. From an image analysis perspective, color variations can occur both within and across tissue sections. For example, within a tissue section, some nuclei may have low chromatin content (e.g., light blue signals), while others may have higher signals (e.g., dark blue); nuclear intensity in one tissue section may be very close to the background intensity (e.g., cytoplasmic, macromolecular components) in another tissue section.

In this paper, we aim to classify components of each his-

*This work was supported by NIH U24 CA1437991 and NIH R01 CA140663 carried out at Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231

tology section in terms of distinct phenotypes (e.g., tumor, stroma, necrosis). We suggest that, compared with human engineered features, unsupervised feature learning is more tolerant to batch effect (e.g., technical variations associated with sample preparation) and can learn pertinent features without user intervention. The key concept is that stacked predictive sparse decomposition (PSD) [23] can elucidate a superior representation that captures intrinsic phenotypic signature. When this representation is coupled with spatial pyramid matching (SPM) [26], which utilizes sparse tissue morphometric signatures at various locations and scales, an improved classification performance is realized.

Organization of this paper is as follows: Section 2 reviews related works. Section 3 describes the details of our proposed approach. Section 4 elaborates the details of our experimental setup, followed by a detailed discussion on the experimental results. Lastly, section 5 concludes the paper.

2. Related Work

Several outstanding reviews for the histology sections analysis can be found in [14, 20]. From our perspective, four distinct works have defined the trends in tissue histology analysis: (i) one group of researchers proposed nuclear segmentation and organization for tumor grading and/or the prediction of tumor recurrence [2, 13, 3, 15, 10]. (ii) A second group of researchers focused on patch level analysis (e.g., small regions) [4, 24, 21], using color and texture features, for tumor representation. (iii) A third group focused on block-level analysis to distinguish different states of tissue development using cell-graph representation [1, 5]. (iv) Finally, a fourth group has suggested detection and representation of the auto-immune response as a prognostic tool for cancer [19].

The major challenge for tissue classification is the large amounts of technical variations and biological heterogeneities in the data [25], which typically results in techniques that are tumor type specific. To overcome this problem, recent studies have focused on either fine tuning human engineered features [4, 24, 25], or applying automatic feature learning [22] for robust representation.

In the context of computer vision research on image categorization, the traditional bag of features (BoF) model has been widely studied and improved through different variations [8, 7, 16, 31, 26], among which SPM [26] has clearly become the major component of the state-of-art systems [17] for its effectiveness in practice.

The evolution of our research in patch level analysis has been SIFT-like feature extraction followed by a evaluation of several kernel-based classification policies [21]; independent subspace analysis that utilizes unsupervised learning without the constraint of being able to reconstruct the original signal [27]; a single layer predictive sparse coding with SVM classifier [32]; and more recently, coupling of either

prior knowledge [9] or predictive sparse coding [11] with with spatial pyramid matching. The current research builds on these results to render an unsupervised feature learning approach with superior performance.

In summary, motivated by the fact that (i) pathologists often use “context” to assess the disease state, (ii) SPM partially captures context [26, 23], and (iii) unsupervised feature learning is preferable to capture the variance in large cohorts, we have extended our previous research with improved scalability and performance.

3. Approach

In this work (PSDⁿSPM), we employ predictive sparse decomposition (PSD) [23] as a building block for the purpose of constructing hierarchical learning framework, which can capture higher-level sparse tissue morphometric features [34]. Unlike many unsupervised feature learning algorithms [28, 29, 33, 39], the feed-forward feature inference of PSD is very efficient, as it involves only element-wise nonlinearity and matrix multiplication. For classification, the predicted sparse features are used in a similar fashion as SIFT features in the traditional framework of SPM, as shown in Figure 1.

3.1. Unsupervised Feature Learning

Given $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{m \times N}$ as a set of vectorized image patches, we formulate the PSD optimization problem as:

$$\begin{aligned} \min_{\mathbf{B}, \mathbf{Z}, \mathbf{G}, \mathbf{W}} \quad & \|\mathbf{X} - \mathbf{B}\mathbf{Z}\|_F^2 + \lambda \|\mathbf{Z}\|_1 + \|\mathbf{Z} - \mathbf{G}\sigma(\mathbf{W}\mathbf{X})\|_F^2 \\ \text{s.t.} \quad & \|\mathbf{b}_i\|_2^2 = 1, \forall i = 1, \dots, h \end{aligned} \quad (1)$$

where $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_h] \in \mathbb{R}^{m \times h}$ is a set of the basis functions; $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_N] \in \mathbb{R}^{h \times N}$ is the sparse feature matrix; $\mathbf{W} \in \mathbb{R}^{h \times m}$ is the auto-encoder; $\mathbf{G} = \text{diag}(g_1, \dots, g_h) \in \mathbb{R}^{h \times h}$ is a scaling matrix with diag being an operator aligning vector, $[g_1, \dots, g_h]$, along the diagonal; $\sigma(\cdot)$ is the element-wise sigmoid function; and λ is a regularization constant. Joint minimization of Eq. (1) with respect to the quadruple $\langle \mathbf{B}, \mathbf{Z}, \mathbf{G}, \mathbf{W} \rangle$, enforces the inference of the nonlinear regressor $\mathbf{G}\sigma(\mathbf{W}\mathbf{X})$ to be similar to the optimal sparse codes, \mathbf{Z} , which can reconstruct \mathbf{X} over \mathbf{B} [23].

As shown below, optimization of Eq. (1) is iterative, where the algorithm terminates when either the objective function is below a preset threshold or the maximum number of iterations has been reached.

1. Randomly initialize \mathbf{B} , \mathbf{W} , and \mathbf{G} .
2. Fixing \mathbf{B} , \mathbf{W} and \mathbf{G} , minimize Eq. (1) with respect to \mathbf{Z} , where \mathbf{Z} can be either solved as a ℓ_1 -minimization problem [28] or equivalently solved by greedy algorithms, e.g., Orthogonal Matching Pursuit (OMP) [35].

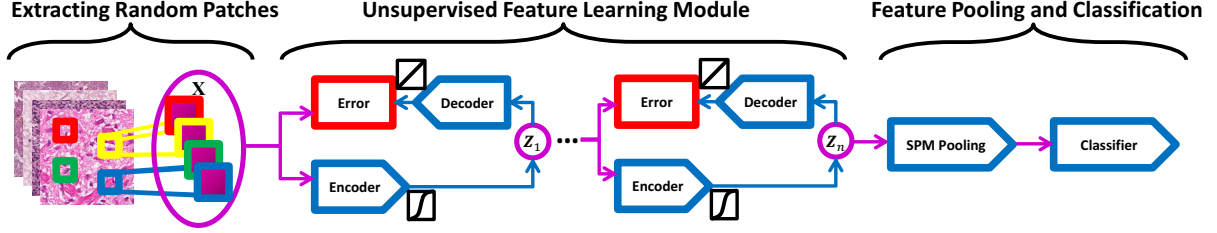


Figure 1. Computational workflow of our approach (PSD^rSPM).

3. Fixing \mathbf{B} , \mathbf{W} and \mathbf{Z} , solve for \mathbf{G} , which is a simple least-square problem with analytic solution.
4. Fixing \mathbf{Z} and \mathbf{G} , update \mathbf{B} and \mathbf{W} , respectively, using the stochastic gradient descent algorithm.
5. Repeat [2]-[4] until stopping condition is satisfied.

In large-scale feature learning problems, involving $\sim 10^5$ image patches, it is computationally intensive to evaluate the sum-gradient over the entire training set. However, both stochastic gradient descent algorithm and GPU parallel computing can provide a significant increase in speed. The former approximates the true gradient of the objective function by the gradient evaluated over mini-batches, and the latter further accelerates the process (up to 5X) with our Matlab implementation based on an Nvidia GTX 580 graphics card. Figure 2 illustrates 1024 basis functions computed from the GBM dataset, which capture both color and texture information from the data and is generally difficult to realize using hand-engineered features.



Figure 2. Computed basis functions (\mathbf{B}) from the Glioblastoma Multiforme (GBM) dataset.

3.2. Spatial Pyramid Matching (SPM)

Having computed the sparse features, $\mathbf{Z} \in \mathbb{R}^{h \times N}$ (e.g., predictions by the nonlinear regressor $\mathbf{G}\sigma(\mathbf{W}\mathbf{X})$), we then construct a code book and proceed with SPM pooling.

The codebook, $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_K] \in \mathbb{R}^{h \times K}$, consisting of K sparse tissue morphometric types, is constructed by solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{C}} \quad & \sum_{i=1}^N \|\mathbf{z}_i - \mathbf{D}\mathbf{c}_i\|^2 \\ \text{s.t.} \quad & \text{card}(\mathbf{c}_i) = 1, \|\mathbf{c}_i\|_1 = 1, \mathbf{c}_i \succeq 0, \forall i \end{aligned} \quad (2)$$

where $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_N] \in \mathbb{R}^{K \times N}$ is the code matrix assigning each \mathbf{z}_i to its closest sparse tissue morphometric type in \mathbf{D} , $\text{card}(\mathbf{c}_i)$ is a cardinality constraint enforcing only one nonzero element in \mathbf{c}_i , and $\mathbf{c}_i \succeq 0$ is a non-negative constraint on all vector elements. Eq. (2) is optimized by alternating between the two variables, *i.e.*, minimizing one while keeping the other fixed. After training, \mathbf{D} is fixed and the query signal set, \mathbf{Z} , is encoded by solving Eq. (2) with respect to \mathbf{C} only.

The next step is to construct a spatial histogram for SPM [26]. By repeatedly subdividing an image, histograms of different sparse tissue morphometric types over the resulting subregions are computed. The spatial histogram, H , is then formed by concatenating the appropriately weighted histograms of sparse tissue morphometric types at all resolutions, *i.e.*,

$$\begin{aligned} H_0 &= H_0^0 \\ H_l &= (H_l^1, \dots, H_l^L), 1 \leq l \leq L \\ H &= \left(\frac{1}{2^L} H_0, \frac{1}{2^L} H_1, \dots, \frac{1}{2^{L-l+1}} H_l, \dots, \frac{1}{2} H_L \right) \end{aligned} \quad (3)$$

where (\cdot) denotes the vector concatenation operator, $l \in \{0, \dots, L\}$ is the resolution level of the image pyramid, and H_l represents the concatenation of histograms for all image subregions at pyramid level l . Instead of using kernel SVM, we employ the homogeneous kernel map [36] and linear SVM [18] for improved efficiency.

4. Experiments And Discussion

In this section, we provide details of the experimental design that includes data from tumor histopathology and 3D cell culture assays. The tumor data includes a curated set of Glioblastoma Multiforme (GBM) and Kidney Clear Cell Carcinoma (KIRC) from The Cancer Genome Atlas (TCGA), which are publicly available from the NIH repository. The 3D cell culture data consists of a tumorigenic breast cancer and a control cell line.

4.1. Classification of Tumor Histopathology

We have evaluated the proposed method against four other techniques for classification of distinct histopathology for two tumor types.

1. PSDⁿSPM^{NR}: The nonlinear kernel SPM that uses spatial-pyramid histograms of sparse tissue morphometric types. In this implementation,
 - (a) $n = 1, 2$;
 - (b) The nonlinear regressor ($\mathbf{Z} = \mathbf{G}\sigma(\mathbf{WX})$) was trained for the inference of \mathbf{Z} ;
 - (c) The image patch size is fixed to be 20×20 and the number of basis functions in the top layer was fixed to be 1024. We adopted the SPAMS optimization toolbox [30] for efficient implementation of OMP to compute the sparse code, \mathbf{Z} , with sparsity prior set to 30;
 - (d) Standard K-means clustering was used for the construction of the dictionary;
 - (e) The level of pyramid was fixed to be 3; and
 - (f) The homogeneous kernel map was applied, followed by the linear SVM for classification.
2. PSD¹SPM^{LR} [11]: The nonlinear kernel SPM that uses spatial-pyramid histograms of sparse tissue morphometric types. In this implementation,
 - (a) The linear regressor ($\mathbf{Z} = \mathbf{WX}$) was trained for the inference of \mathbf{Z} ;
 - (b) For consistency, the image patch size and the number of basis functions was fixed at 20×20 and 1024, respectively. The sparsity constraint was set at 0.3 for best performance following cross validation.
 - (c) Standard K-means clustering was used for the construction of the dictionary;
 - (d) The level of pyramid was fixed to be 3;
 - (e) The homogeneous kernel map was applied, followed by linear SVM for classification.
3. ScSPM [37]: The linear SPM that utilizes linear kernel on spatial-pyramid pooling of SIFT sparse codes. In this implementation,
 - (a) The dense SIFT features was extracted on 16×16 patches sampled from each image on a grid with stepsize 8 pixels;
 - (b) The sparse constraint parameter λ was fixed to be 0.15, which was determined empirically to achieve the best performance;
 - (c) The level of pyramid was fixed to be 3;
 - (d) Linear SVM was used for classification.
4. KSPM [26]: The nonlinear kernel SPM that uses spatial-pyramid histograms of SIFT features; In the implementation,
 - (a) The dense SIFT features was extracted on 16×16 patches sampled from each image on a grid with stepsize 8 pixels;
 - (b) Standard K-means clustering was used for the construction of the dictionary;
 - (c) The level of pyramid was fixed to be 3;
 - (d) The homogeneous kernel map was applied, followed by linear SVM for classification.
5. CTSPM: The nonlinear kernel SPM that uses spatial-pyramid histograms of color and texture features; In this implementation,
 - (a) Color features were extracted from the RGB color space;
 - (b) Texture features were extracted via steerable filters [38] with 4 directions ($\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$) and 5 scales ($\sigma \in \{1, 2, 3, 4, 5\}$) from the grayscale image;
 - (c) The feature vector was constructed by concatenating texture and mean color on 20×20 patches, empirically, to achieve the best performance;
 - (d) Standard K-means clustering was used for the construction of the dictionary;
 - (e) The level of pyramid was fixed to be 3;
 - (f) The homogeneous kernel map was applied, followed by linear SVM for classification.

All experimental processes were repeated 10 times with randomly selected training and testing images. The final results were reported as the mean and standard deviation of the classification rates on the following two distinct datasets, which included vastly different tumor types:

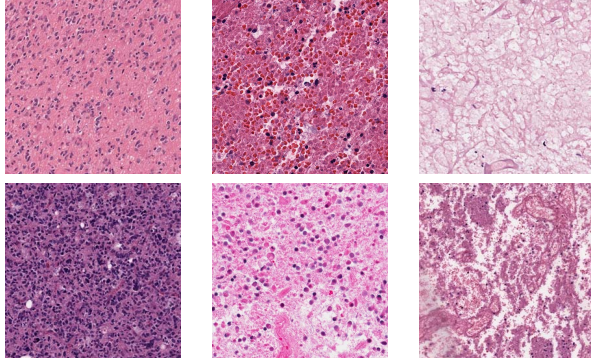


Figure 3. GBM Examples. First column: Tumor; Second column: Transition to necrosis; Third column: Necrosis.

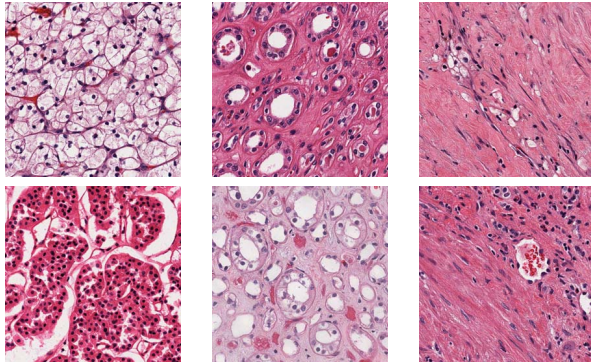


Figure 4. KIRC examples. First column: Tumor; Second column: Normal; Third column: Stromal.

1. **GBM Dataset.** The GBM dataset contains 3 classes: Tumor, Necrosis, and Transition to Necrosis, which were curated from whole slide images (WSI) scanned with a 20X objective (0.502 micron/pixel). Examples can be found in Figure 3. The number of images per category are 628, 428 and 324, respectively. Most images are 1000×1000 pixels. In this experiment, we trained on 40, 80 and 160 images per category and tested on the rest, using three different dictionary sizes: 256, 512 and 1024. Detailed comparisons are shown in Table 1.
2. **KIRC Dataset.** The KIRC dataset contains 3 classes: Tumor, Normal, and Stromal, which were curated from whole slide images (WSI) scanned with a 40X objective (0.252 micron/pixel). Examples can be found in Figure 4. The number of images per category are 568, 796 and 784, respectively. Most images are 1000×1000 pixels. In this experiment, we trained on 70, 140 and 280 images per category and tested on the rest, using three different dictionary sizes: 256, 512 and 1024. Detailed comparisons are shown in Table 2.

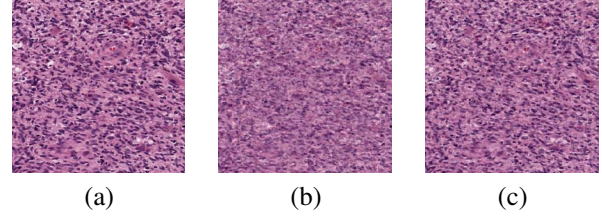


Figure 5. Comparison of PSD with linear and nonlinear regressors in terms of reconstruction. (a) Original image; (b) Reconstruction by PSD with linear regressor (SNR=14.9429); (c) Reconstruction by PSD with nonlinear regressor (SNR=19.3436).

4.2. Discussion

Above experiments indicate that,

1. Features from unsupervised feature learning are more tolerant to the batch effect than human engineered features for tissue classification. Tables 1 and 2 show that PSDⁿSPM consistently outperforms KSPM, ScSPM and CTSPM on the two distinct datasets that suffer from technical variations as a result of both sample preparation and biological heterogeneity, where the latter is due to the variation in tumor phenotype across patients.
2. PSD with nonlinear regressor outperforms PSD with linear regressor in terms of both reconstruction and classification, as shown in Figure 5 as well as Tables 1 and 2.
3. Stacking multiple layers of PSD enables learning higher level features, which further improves the classification performance.

As a result, the proposed approach has the following merits,

1. **Extensibility to different tumor types.** Tables 1 and 2 confirm the superiority and consistency in the performance of the proposed approach on two vastly different tumor types, which are due to the improved generalization ability of features from unsupervised feature learning, compared to human engineered features (e.g., SIFT), and ultimately ensures the extensibility of proposed approach to different tumor types.
2. **Robustness in the presence of large amounts of technical variations and biological heterogeneities.** Tables 1 and 2 indicate that the performance of our approach, based on small number of training samples, is comparable to or better than the performance of ScSPM, KSPM and CTSPM, which are based on large number of training samples. Given the fact that TCGA datasets contain large amounts of technical variations and biological heterogeneities, these results clearly indicate the robustness of our approach, which improves the

	Method	DictionarySize=256	DictionarySize=512	DictionarySize=1024
160 training	PSD ² SPM ^{NR}	91.85 ± 1.03	91.86 ± 0.78	92.07 ± 0.65
	PSD ¹ SPM ^{NR}	91.85 ± 0.69	91.89 ± 0.99	91.74 ± 0.85
	PSD ¹ SPM ^{LR} [11]	91.02 ± 1.89	91.41 ± 0.95	91.20 ± 1.29
	ScSPM [37]	79.58 ± 0.61	81.29 ± 0.86	82.36 ± 1.10
	KSPM [26]	85.00 ± 0.79	86.47 ± 0.55	86.81 ± 0.45
	CTSPM	78.61 ± 1.33	78.71 ± 1.18	78.69 ± 0.81
80 training	PSD ² SPM ^{NR}	90.51 ± 1.06	90.88 ± 0.66	90.51 ± 1.06
	PSD ¹ SPM ^{NR}	90.74 ± 0.95	90.42 ± 0.94	89.70 ± 1.20
	PSD ¹ SPM ^{LR} [11]	88.63 ± 0.91	88.91 ± 1.18	88.64 ± 1.08
	ScSPM [37]	77.65 ± 1.43	78.31 ± 1.13	81.00 ± 0.98
	KSPM [26]	83.81 ± 1.22	84.32 ± 0.67	84.49 ± 0.34
	CTSPM	75.93 ± 1.18	76.06 ± 1.52	76.19 ± 1.33
40 training	PSD ² SPM ^{NR}	87.90 ± 0.91	88.21 ± 0.90	87.71 ± 0.81
	PSD ¹ SPM ^{NR}	87.72 ± 1.21	86.99 ± 1.76	86.33 ± 1.32
	PSD ¹ SPM ^{LR} [11]	84.06 ± 1.16	83.72 ± 1.46	83.40 ± 1.14
	ScSPM [37]	73.60 ± 1.68	75.58 ± 1.29	76.24 ± 3.05
	KSPM [26]	80.54 ± 1.21	80.56 ± 1.24	80.46 ± 0.56
	CTSPM	73.10 ± 1.51	72.90 ± 1.09	72.65 ± 1.41

Table 1. Performance of different methods on the GBM dataset.

	Method	DictionarySize=256	DictionarySize=512	DictionarySize=1024
280 training	PSD ² SPM ^{NR}	99.03 ± 0.20	98.89 ± 0.19	98.92 ± 0.21
	PSD ¹ SPM ^{NR}	98.98 ± 0.35	98.81 ± 0.45	98.69 ± 0.41
	PSD ¹ SPM ^{LR} [11]	97.19 ± 0.49	97.27 ± 0.44	97.08 ± 0.45
	ScSPM [37]	94.52 ± 0.44	96.37 ± 0.45	96.81 ± 0.50
	KSPM [26]	93.55 ± 0.31	93.76 ± 0.27	93.90 ± 0.19
	CTSPM	87.45 ± 0.59	87.95 ± 0.49	88.53 ± 0.49
140 training	PSD ² SPM ^{NR}	98.26 ± 0.34	98.07 ± 0.46	97.85 ± 0.56
	PSD ¹ SPM ^{NR}	98.17 ± 0.72	98.05 ± 0.71	97.99 ± 0.82
	PSD ¹ SPM ^{LR} [11]	96.80 ± 0.75	96.52 ± 0.76	96.55 ± 0.84
	ScSPM [37]	93.46 ± 0.55	95.68 ± 0.36	96.76 ± 0.63
	KSPM [26]	92.50 ± 1.12	93.06 ± 0.82	93.26 ± 0.68
	CTSPM	86.55 ± 0.99	86.40 ± 0.54	86.49 ± 0.58
70 training	PSD ² SPM ^{NR}	96.67 ± 0.53	96.20 ± 0.54	95.57 ± 0.66
	PSD ¹ SPM ^{NR}	96.42 ± 0.68	96.41 ± 0.59	96.03 ± 0.69
	PSD ¹ SPM ^{LR} [11]	95.12 ± 0.54	95.13 ± 0.51	95.09 ± 0.40
	ScSPM [37]	91.93 ± 1.00	93.67 ± 0.72	94.86 ± 0.86
	KSPM [26]	90.78 ± 0.98	91.34 ± 1.13	91.59 ± 0.97
	CTSPM	84.76 ± 1.32	84.29 ± 1.53	83.71 ± 1.42

Table 2. Performance of different methods on the KIRC dataset.

scalability with varying training sample sizes, and the reliability of further analysis on large cohort of whole mount tissue sections.

4.3. Application to 3D Cell Culture Models

To demonstrate extensibility of the proposed method, the technique has been applied to discriminated colony organization in 3D cell culture models for control (MCF10A) and malignant (MDA-MB-231) cell lines, which have been

well characterized in [6]. Five biological replicates per cell line were harvested on Day 5 and imaged with confocal microscopy with the middle sections shown in Figure 6. All images were scaled isotropically with the image patch size fixed to be $20 \times 20 \times 20$ in the isotropic space for unsupervised feature learning. The performance is reported as the mean classification rate (shown in Table 3), which clearly indicates the extensibility of proposed method for classification of aberrant growth conditions.

	Method	DictionarySize=256	DictionarySize=512	DictionarySize=1024
2 training	PSD ¹ SPM ^{NR}	61.67	81.67	88.33

Table 3. Performance of proposed method on 3D cell culture differentiation.

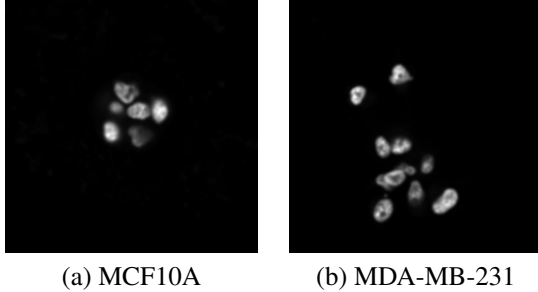


Figure 6. Examples of a section for the control and malignant lines.

5. Conclusion and Future Work

In this paper, we proposed a multi-layer PSD framework for classification of distinct regions of tumor histopathology, which outperforms traditional methods that are typically based on pixel- or patch-level features. Our analysis indicates that the proposed approach is (i) extensible to different tumor types; (ii) robust in the presence of large amounts of technical variations and biological heterogeneities; (iii) scalable with varying training sample sizes; and (iv) extensible to 3D cell culture models. Future research will focus on further validation of our approach on other tumor types and further discrimination of phenotypic responses in multicellular systems.

6. Disclaimer

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

References

- [1] E. Acar, G. Plopper, and B. Yener. Coupled analysis of in vitro and histology samples to quantify structure-function relationships. *PLoS One*, 7(3):e32227, 2012. 2
- [2] D. Axelrod, N. Miller, H. Lickley, J. Qian, W. Christens-Barry, Y. Yuan, Y. Fu, and J. Chapman. Effect of quantitative nuclear features on recurrence of ductal carcinoma in situ (DCIS) of breast. *Cancer Informatics*, 4:99–109, 2008. 2
- [3] A. Basavanthally, J. Xu, A. Madabhushu, and S. Ganesan. Computer-aided prognosis of ER+ breast cancer histopathology and correlating survival outcome with oncotype DX assay. In *ISBI*, pages 851–854, 2009. 2
- [4] R. Bhagavatula, M. Fickus, W. Kelly, C. Guo, J. Ozolek, C. Castro, and J. Kovacevic. Automatic identification and delineation of germ layer components in *h&e* stained images of teratomas derived from human and nonhuman primate embryonic stem cells. In *ISBI*, pages 1041–1044, 2010. 2
- [5] C. Bilgin, S. Ray, B. Baydil, W. Daley, M. Larsen, and B. Yener. Multiscale feature analysis of salivary gland branching morphogenesis. *PLoS One*, 7(3):e32906, 2012. 2
- [6] C. C. Bilgin, S. Kim, E. leung, H. Chang, and B. Parvin. Integrated profiling of three dimensional cell culture models and 3d microscopy. *Bioinformatics*, 2013. 6
- [7] O. Boiman, E. Shechtman, and M. Irani. In defense of nearest-neighbor based image classification. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008. 2
- [8] A. Bosch, A. Zisserman, and X. Muñoz. Scene classification using a hybrid generative/discriminative approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(4):712–727, Apr. 2008. 2
- [9] H. Chang, A. Borowsky, P. Spellman, and B. Parvin. Classification of tumor histology via morphometric context. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2013. 2
- [10] H. Chang, J. Han, A. Borowsky, L. A. Loss, J. W. Gray, P. T. Spellman, and B. Parvin. Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association. *IEEE Trans. Med. Imaging*, 32(4):670–682, 2013. 2
- [11] H. Chang, N. Nayak, P. Spellman, and B. Parvin. Characterization of tissue histopathology via predictive sparse decomposition and spatial pyramid matching. *Medical image computing and computed-assisted intervention—MICCAI*, 2013. 2, 4, 6
- [12] L. Dalton, S. Pinder, C. Elston, I. Ellis, D. Page, W. Dupont, and R. Blamey. Histological gradings of breast cancer: linkage of patient outcome with level of pathologist agreements. *Modern Pathology*, 13(7):730–735, 2000. 1

- [13] M. Datar, D. Padfield, and H. Cline. Color and texture based segmentation of molecular pathology images using HSOMs. In *ISBI*, pages 292–295, 2008. 2
- [14] C. Demir and B. Yener. Automated cancer diagnosis based on histopathological images: A systematic survey. *Technical Report, Rensselaer Polytechnic Institute, Department of Computer Science.*, 2009. 2
- [15] S. Doyle, M. Feldman, J. Tomaszewski, N. Shih, and A. Madabhushu. Cascaded multi-class pairwise classifier (CASCAMPA) for normal, cancerous, and cancer confounder classes in prostate histology. In *ISBI*, pages 715–718, 2011. 2
- [16] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, Dec. 2006. 2
- [17] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>. 2
- [18] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008. 3
- [19] H. Fatakdawala, J. Xu, A. Basavanthally, G. Bhanot, S. Ganesan, F. Feldman, J. Tomaszewski, and A. Madabhushi. Expectation-maximization-driven geodesic active contours with overlap resolution (EMaGACOR): Application to lymphocyte segmentation on breast cancer histopathology. *IEEE Transactions on Biomedical Engineering*, 57(7):1676–1690, 2010. 2
- [20] M. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, and Y. Bulent. Histopathological image analysis: a review. *IEEE Transactions on Biomedical Engineering*, 2:147–171, 2009. 2
- [21] J. Han, H. Chang, L. Loss, K. Zhang, F. Baehner, J. Gray, P. Spellman, and B. Parvin. Comparison of sparse coding and kernel methods for histopathological classification of glioblastoma multiforme. In *ISBI*, pages 711–714, 2011. 2
- [22] C. Huang, A. Veillard, N. Lomeine, D. Racoceanu, and L. Roux. Time efficient sparse analysis of histopathological whole slide images. *Computerized medical imaging and graphics*, 35(7-8):579–591, 2011. 2
- [23] K. Kavukcuoglu, M. Ranzato, and Y. LeCun. Fast inference in sparse coding algorithms with applications to object recognition. Technical Report CBLL-TR-2008-12-01, Computational and Biological Learning Lab, Courant Institute, NYU, 2008. 2
- [24] J. Kong, L. Cooper, A. Sharma, T. Kurk, D. Brat, and J. Saltz. Texture based image recognition in microscopy images of diffuse gliomas with multi-class gentle boosting mechanism. In *ICASSAP*, pages 457–460, 2010. 2
- [25] S. Kothari, J. Phan, A. Osunkoya, and M. Wang. Biological interpretation of morphological patterns in histopathological whole slide images. In *ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, 2012. 2
- [26] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 2169–2178, 2006. 2, 3, 4, 6
- [27] Q. Le, J. Han, J. Gray, P. Spellman, A. Borowsky, and B. Parvin. Learning invariant features from tumor signature. In *ISBI*, pages 302–305, 2012. 2
- [28] H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In *In NIPS*, pages 801–808. NIPS, 2007. 2
- [29] H. Lee, C. Ekanadham, and A. Y. Ng. Sparse deep belief net model for visual area v2. In *Advances in Neural Information Processing Systems 20*. MIT Press, 2008. 2
- [30] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res.*, 11:19–60, Mar. 2010. 4
- [31] F. Moosmann, B. Triggs, and F. Jurie. Randomized clustering forests for building fast and discriminative visual vocabularies. In *NIPS*, 2006. 2
- [32] N. Nayak, H. Chang, A. Borowsky, P. Spellman, and B. Parvin. Classification of tumor histopathology via sparse feature learning. In *Proc. ISBI*, pages 410–413, 2013. 2
- [33] C. Poultney, S. Chopra, and Y. Lecun. Efficient learning of sparse representations with an energy-based model. In *Advances in Neural Information Processing Systems (NIPS 2006)*. MIT Press, 2006. 2
- [34] M. Ranzato, Y.-L. Boureau, and Y. LeCun. Sparse feature learning for deep belief networks. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 1185–1192. MIT Press, Cambridge, MA, 2008. 2
- [35] J. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 53(12):4655–4666, 2007. 2
- [36] A. Vedaldi and A. Zisserman. Efficient additive kernels via explicit feature maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):480–492, 2012. 3
- [37] J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 1794–1801, 2009. 4, 6
- [38] R. A. Young and R. M. Lesperance. The gaussian derivative model for spatial-temporal vision. *I. Cortical Model. Spatial Vision*, 2001:3–4, 2001. 4
- [39] K. Yu, T. Zhang, and Y. Gong. Nonlinear learning using local coordinate coding. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 2223–2231, 2009. 2