

Published in final edited form as:

Proc IEEE Int Conf Comput Vis. 2013 ; : 169–176. doi:10.1109/ICCV.2013.28.

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

Hang Chang^{1,*,\dagger}, Yin Zhou^{1,\dagger}, Paul Spellman², and Bahram Parvin^{1,\dagger}

Hang Chang: hchang@lbl.gov; Yin Zhou: yinzhou@lbl.gov; Paul Spellman: spellmap@ohsu.edu; Bahram Parvin: b_parvin@lbl.gov

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

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

Abstract

Image-based classification of tissue histology, in terms of distinct histopathology (e.g., tumor or necrosis regions), 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 clinical outcome. However, the performance of the existing techniques is hindered as a result of large technical variations (e.g., fixation, staining) and biological heterogeneities (e.g., cell type, cell state) that are always present in a large cohort. We suggest that, compared with human engineered features widely adopted in existing systems, unsupervised feature learning is more tolerant to batch effect (e.g., technical variations associated with sample preparation) and pertinent features can be learned without user intervention. This leads to a novel approach for classification of tissue histology based on unsupervised feature learning and spatial pyramid matching (SPM), which utilize sparse tissue morphometric signatures at various locations and scales. This approach has been evaluated on two distinct datasets consisting of different tumor types collected from The Cancer Genome Atlas (TCGA), and the experimental results indicate that the proposed approach is (i) extensible to different tumor types; (ii) robust in the presence of wide technical variations and biological heterogeneities; and (iii) scalable with varying training sample sizes.

1. Introduction

Tumor histology provides a detailed insight into cellular morphology, organization, and heterogeneity. For example, tumor histological sections can be used to identify mitotic cells, cellular aneuploidy, and autoimmune responses. More importantly, if tumor morphology and architecture can be quantified on large histological datasets, then it will pave the way for constructing histological databases that are prognostic, the same way that genome analysis techniques have identified molecular subtypes and predictive markers. Genome wide analysis techniques (e.g., microarray analysis) have the advantages of standardized tools for

*This work was supported by 1) NIH U24 CA1437991 carried out at Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231; and 2) NIH R01 CA140663 carried out at Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231

^{*}Co-Corresponding authors.

[†]Co-First authors

data analysis and pathway enrichment, which enables hypothesis generation for the underlying mechanism. On the other hand, histological signatures are hard to compute because of the technical variations and biological heterogeneities in the stained histological sections; however, they offer insights into tissue composition as well as heterogeneity (e.g., mixed populations) and rare events.

Histological sections are often stained with hematoxylin and eosin stains (H&E), which label DNA and protein contents, respectively. Traditional histological analysis is performed by a trained pathologist through the characterization of phenotypic content, such as various cell types, cellular organization, cell state and health, and cellular secretion. While, such manual analysis may incur inter- and intra-observer variations [10]. The value of the quantitative histological image analysis originates from its capability in capturing detailed morphometric features on a cell-by-cell basis and the organization of cells. Such rich descriptions can then be linked with genomic information and survival distribution as an improved basis for diagnosis and therapy. Additionally, in the presence of large datasets, quantitative histological signatures can be used to identify intrinsic subtypes of a specific tumor type, which is supplementary to histological tumor grading.

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 quantify composition of each tissue section in terms of distinct histopathology, such as tumor or necrosis regions. 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. Here, we propose a tissue classification method based on unsupervised feature learning and spatial pyramid matching (SPM) [24], which utilize sparse tissue morphometric signatures at various locations and scales.

Due to the robustness of unsupervised feature learning and the effectiveness of the spatial pyramid matching framework, our approach achieves excellent performance even with a small number of training samples across independent datasets of vastly different tumor types.

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 detailed discussion on the experimental results. Lastly, section 5 concludes the paper.

2. Related Work

Several outstanding reviews for the analysis of histology sections can be found in [12, 18]. 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, 11, 3, 13]. (ii) A second group of researchers focused on patch level analysis (e.g., small regions) [4, 22, 19], 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 [17].

The major challenge for tissue classification is the large amounts of technical variations and biological heterogeneities in the data [23], 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, 22, 23], or applying automatic feature learning [20], for robust representation.

In the context of image categorization research, the traditional bag of features (BoF) model has been widely studied and improved through different variations, i.e., modeling of co-occurrence of descriptors based on generative methods [8, 7, 27, 31], improving dictionary construction through discriminative learning [14, 29], modeling the spatial layout of local descriptors based on SPM kernel [24]. It is clear that SPM has become the major component of the state-of-art systems [15] for its effectiveness in practice.

Pathologists often use “context” to assess the disease state, and SPM partially captures context as a result of its hierarchical nature. Motivated by the works of [24, 21], we have therefore encoded sparse tissue morphometric signatures, at different locations and scales within the SPM framework. The end results are highly robust and effective systems across multiple tumor types utilizing a limited number of training samples.

3. Approach

In this work (PSDⁿSPM), we employ the predictive sparse decomposition (PSD) [21] as the building block for the purpose of constructing hierarchical learning framework, which can capture higher-level sparse tissue morphometric features. Unlike many unsupervised feature learning algorithms [25, 26, 30, 36], 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:

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

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. The goal of jointly minimizing Eq. (1) with respect to the quadruple $\langle \mathbf{B}, \mathbf{Z}, \mathbf{G}, \mathbf{W} \rangle$ is to enforce the inference of the nonlinear regressor $\mathbf{G}\sigma(\mathbf{W}\mathbf{X})$ to be resemble to the optimal sparse codes \mathbf{Z} that can reconstruct \mathbf{X} over \mathbf{B} [21].

An iterative process is employed for optimizing Eq. (1), as detailed below. The algorithm is terminated once either the objective function is below a preset threshold or a maximum number of iterations is 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 [25] or equivalently solved by greedy algorithms, e.g., Orthogonal Matching Pursuit (OMP) [32].
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. Therefore, to improve the scalability of our approach, we use both stochastic gradient descent algorithm and GPU parallel computing (based on an Nvidia GTX 580 graphics card). 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. 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.

3.2. Spatial Pyramid Matching (SPM)

Now suppose we have obtained the sparse features $\mathbf{Z} \in \mathbb{R}^{h \times N}$, *i.e.*, predictions by the nonlinear regressor $\mathbf{G}\sigma(\mathbf{W}\mathbf{X})$. We first construct a codebook $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_K] \in \mathbb{R}^{h \times K}$, which consists of K sparse tissue morphometric types, by solving the 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 , $\mathbf{c}_i \succeq 0$ 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 second step is to construct spatial histogram for SPM [24]. By repeatedly subdividing an image, we compute the histograms of different sparse tissue morphometric types over the resulting subregions. Then, the spatial histogram, H , is formed by concatenating the appropriately weighted histograms of sparse tissue morphometric types at all resolutions, *i.e.*,

$$\begin{aligned} H_0 &= H_0^0 \\ H_l &= (H_l^1, \dots, H_l^4), 1 \leq l \leq L \\ H &= (\frac{1}{2^L} H_0, \frac{1}{2^L} H_1, \dots, \frac{1}{2^{L-l+1}} H_l, \dots, \frac{1}{2} H_L) \end{aligned} \quad (3)$$

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 [33] and linear SVM [16] for improved efficiency.

4. Experiments And Discussion

In this section, we give the details of the datasets and methods involved in our experiments, which are followed by a thorough discussion on the experimental results.

4.1. Experimental Setup

We have evaluated five classification methods on two distinct datasets, curated from (i) Glioblastoma Multiforme (GBM) and (ii) Kidney Renal Clear Cell Carcinoma (KIRC) from The Cancer Genome Atlas (TCGA), which are publicly available from the NIH (National Institute of Health) repository. The five methods and detailed implementations are:

1. PSDⁿSPM^{NR}: the nonlinear kernel SPM that uses spatial-pyramid histograms of sparse tissue morphometric types. In the implementation,
 - a. $n = 1, 2$;
 - b. nonlinear regressor ($\mathbf{Z} = \mathbf{G}\sigma(\mathbf{W}\mathbf{X})$) is trained for the inference of \mathbf{Z} ;
 - c. the image patch size was fixed to be 20×20 and the number of basis functions in the top layer was fixed to be 1024. We adopt the SPAMS

optimization toolbox [28] 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;
 - f. homogeneous kernel map was applied, followed by linear SVM for classification.
2. PSD¹SPM^{LR} [9]: the nonlinear kernel SPM that uses spatial-pyramid histograms of sparse tissue morphometric types. In the implementation,
 - a. linear regressor ($\mathbf{Z} = \mathbf{W}\mathbf{X}$) is trained for the inference of \mathbf{Z} ;
 - b. for fair comparison, the image patch size was fixed to be 20×20, and the number of basis functions was fixed to be 1024. To achieve the best performance, the sparse constraint parameter λ was fixed to be 0.3.
 - c. standard K-means clustering was used for the construction of the dictionary;
 - d. the level of pyramid was fixed to be 3;
 - e. homogeneous kernel map was applied, followed by linear SVM for classification.
 3. ScSPM [34]: the linear SPM that uses linear kernel on spatial-pyramid pooling of SIFT sparse codes. 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. the sparse constraint parameter λ was fixed to be 0.15, 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 [24]: 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. 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 the implementation,
 - a. color features were extracted from the RGB color space;
 - b. texture features were extracted via steerable filters [35] with 4 directions ($\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{2}\}$) 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. 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, and 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:

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, with 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, with three different dictionary sizes: 256, 512 and 1024. Detailed comparisons are shown in Table 2.

4.2. Discussion

The experiments, conducted above, indicate that,

1. Features from unsupervised feature learning are more tolerant to batch effect than human engineered features for tissue classification. In our experiments, we show that (see Tables 1 and 2), PSDⁿSPM consistently outperforms KSPM and CTSPM on the two distinct datasets, which suggest that, given the large amounts of technical variations and biological heterogeneities in the TCGA datasets, features

from unsupervised feature learning are more tolerant to batch effect than human engineered features for tissue classification.

2. PSD with nonlinear regressor outperforms PSD with linear regressor in terms of both reconstruction and classification, as shown in Figure 5 and Tables 1 and 2.
3. Stacking multiple layers of PSD enables learning higher level features, thus further improves the 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 performance of the proposed approach on two vastly different tumor types, which is due to the improved generalization ability of features from unsupervised feature learning, compared with human engineered features (e.g., SIFT), and 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 either better than or comparable to the performance of Sc-SPM, KSPM and CTSPM based on large number of training samples. Given the fact that TCGA datasets contain large amount of technical variations and biological heterogeneities, these results clearly indicate the robustness of our approach, which improves the scalability with varying training sample sizes, and the reliability of further analysis on large cohort of whole mount tissue sections.

4.3. Extensibility to Different Cell Culture Model

As a further validation, we also applied our approach for the task of 3D cell culture differentiation, where the 3D cell cultures were grown in Matrigel using on-top method for one non-transformed line (MCF10A) and one malignant line (MDA-MB-231), both of which are breast epithelial lines that have been obtained from ATCC [6]. 5 samples per cell line were collected on day 5 with pixel size of 0.25 micron in X,Y and 1 micron in Z dimensions. An example of middle slices of two different cell cultures are demonstrated in Figure 6. All images were scaled isotropically with image patch size fixed to be $20 \times 20 \times 20$ in the isotropic space for unsupervised feature learning, and the performance is reported as the mean classification rate, as shown in Table 3, which indicates the extensibility of our method to different cell culture models.

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. The most encouraging results of this paper are that our 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 different cell culture models. Future work will be focused on further validating our approach on other tissue types and different cell culture models.

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. Acar E, Plopper G, Yener B. Coupled analysis of in vitro and histology samples to quantify structure-function relationships. *PLoS One*. 2012; 7(3):e32227. [PubMed: 22479315]
2. Axelrod D, Miller N, Lickley H, Qian J, Christens-Barry W, Yuan Y, Fu Y, Chapman J. Effect of quantitative nuclear features on recurrence of ductal carcinoma in situ (DCIS) of breast. *Cancer Informatics*. 2008; 4:99–109. [PubMed: 18779878]
3. Basavanthally A, Xu J, Madabhushu A, Ganesan S. Computer-aided prognosis of ER+ breast cancer histopathology and correlating survival outcome with oncotype DX assay. *ISBI*. 2009:851–854.
4. Bhagavatula R, Fickus M, Kelly W, Guo C, Ozolek J, Castro C, Kovacevic J. Automatic identification and delineation of germ layer components in *h&e* stained images of teratomas derived from human and nonhuman primate embryonic stem cells. *ISBI*. 2010:1041–1044.
5. Bilgin C, Ray S, Baydil B, Daley W, Larsen M, Yener B. Multiscale feature analysis of salivary gland branching morphogenesis. *PLoS One*. 2012; 7(3):e32906. [PubMed: 22403724]
6. Bilgin CC, Kim S, leung E, Chang H, Parvin B. Integrated profiling of three dimensional cell culture models and 3d microscopy. *Bioinformatics*.
7. Boiman, O.; Shechtman, E.; Irani, M. In defense of nearest-neighbor based image classification. *Proceedings of the Conference on Computer Vision and Pattern Recognition*; 2008. p. 1-8.
8. Bosch A, Zisserman A, Muñoz X. Scene classification using a hybrid generative/discriminative approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Apr; 2008 30(4):712–727. [PubMed: 18276975]
9. Chang, H.; Nayak, N.; Spellman, P.; Parvin, B. Medical image computing and computed-assisted intervention–MICCAI. 2013. Characterization of tissue histopathology via predictive sparse decomposition and spatial pyramid matching.
10. Dalton L, Pinder S, Elston C, Ellis I, Page D, Dupont W, Blamey R. Histological gradings of breast cancer: linkage of patient outcome with level of pathologist agreements. *Modern Pathology*. 2000; 13(7):730–735. [PubMed: 10912931]
11. Datar M, Padfield D, Cline H. Color and texture based segmentation of molecular pathology images using HSOMs. *ISBI*. 2008:292–295.
12. Demir, C.; Yener, B. Technical Report. Rensselaer Polytechnic Institute, Department of Computer Science; 2009. Automated cancer diagnosis based on histopathological images: A systematic survey.
13. Doyle S, Feldman M, Tomaszewski J, Shih N, Madabhushu A. Cascaded multi-class pairwise classifier (CASCAMPA) for normal, cancerous, and cancer confounder classes in prostate histology. *ISBI*. 2011:715–718.

14. Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*. Dec; 2006 15(12):3736–3745. [PubMed: 17153947]
15. Everingham, M.; Van Gool, L.; Williams, CKI.; Winn, J.; Zisserman, A. The PASCAL Visual Object Classes Challenge. 2012. (VOC2012) Results. <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>
16. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*. 2008; 9:1871–1874.
17. Fatakdawala H, Xu J, Basavanthally A, Bhanot G, Ganesan S, Feldman F, Tomaszewski J, Madabhushi A. Expectation-maximization-driven geodesic active contours with overlap resolution (EMaGACOR): Application to lymphocyte segmentation on breast cancer histopathology. *IEEE Transactions on Biomedical Engineering*. 2010; 57(7):1676–1690. [PubMed: 20172780]
18. Gurcan M, Boucheron L, Can A, Madabhushi A, Rajpoot N, Bulent Y. Histopathological image analysis: a review. *IEEE Transactions on Biomedical Engineering*. 2009; 2:147–171.
19. Han J, Chang H, Loss L, Zhang K, Baehner F, Gray J, Spellman P, Parvin B. Comparison of sparse coding and kernel methods for histopathological classification of glioblastoma multiforme. *ISBI*. 2011:711–714.
20. Huang C, Veillard A, Lomeine N, Racoceanu D, Roux L. Time efficient sparse analysis of histopathological whole slide images. *Computerized medical imaging and graphics*. 2011; 35(7–8):579–591. [PubMed: 21145705]
21. Kavukcuoglu, K.; Ranzato, M.; LeCun, Y. Technical Report CBLL-TR-2008-12-01. Computational and Biological Learning Lab, Courant Institute; NYU: 2008. Fast inference in sparse coding algorithms with applications to object recognition.
22. Kong J, Cooper L, Sharma A, Kurk T, Brat D, Saltz J. Texture based image recognition in microscopy images of diffuse gliomas with multi-class gentle boosting mechanism. *ICASSAP*. 2010:457–460.
23. Kothari, S.; Phan, J.; Osunkoya, A.; Wang, M. Biological interpretation of morphological patterns in histopathological whole slide images. *ACM Conference on Bioinformatics, Computational Biology and Biomedicine*; 2012.
24. Lazebnik, S.; Schmid, C.; Ponce, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *Proceedings of the Conference on Computer Vision and Pattern Recognition*; 2006. p. 2169-2178.
25. Lee, H.; Battle, A.; Raina, R.; Ng, AY. In *NIPS*. NIPS; 2007. Efficient sparse coding algorithms; p. 801-808.
26. Lee, H.; Ekanadham, C.; Ng, AY. *Advances in Neural Information Processing Systems 20*. MIT Press; 2008. Sparse deep belief net model for visual area v2.
27. Li, F-F.; Perona, P. A bayesian hierarchical model for learning natural scene categories. *Proceedings of the Conference on Computer Vision and Pattern Recognition*; Washington, DC, USA. IEEE Computer Society; 2005. p. 524-531.
28. Mairal J, Bach F, Ponce J, Sapiro G. Online learning for matrix factorization and sparse coding. *J Mach Learn Res*. Mar.2010 11:19–60.
29. Moosmann F, Triggs B, Jurie F. Randomized clustering forests for building fast and discriminative visual vocabularies. *NIPS*. 2006
30. Poultney, C.; Chopra, S.; Lecun, Y. *Advances in Neural Information Processing Systems (NIPS 2006)*. MIT Press; 2006. Efficient learning of sparse representations with an energy-based model.
31. Quelhas, P.; Monay, F.; Odobez, J-M.; Gatica-Perez, D.; Tuytelaars, T.; Van Gool, L. Modeling scenes with local descriptors and latent aspects. *Proceedings of the IEEE International Conference on Computer Vision, ICCV '05*; Washington, DC, USA. IEEE Computer Society; 2005. p. 883-890.
32. Tropp J, Gilbert A. Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*. 2007; 53(12):4655–4666.
33. Vedaldi A, Zisserman A. Efficient additive kernels via explicit feature maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2012; 34(3):480–492. [PubMed: 21808094]

34. Yang, J.; Yu, K.; Gong, Y.; Huang, T. Linear spatial pyramid matching using sparse coding for image classification. Proceedings of the Conference on Computer Vision and Pattern Recognition; 2009. p. 1794-1801.
35. Young RA, Lesperance RM. The gaussian derivative model for spatial-temporal vision. I. Cortical Model. Spatial Vision. 2001; 2001:3-4.
36. Yu, K.; Zhang, T.; Gong, Y. Nonlinear learning using local coordinate coding. In: Bengio, Y.; Schuurmans, D.; Lafferty, J.; Williams, CKI.; Culotta, A., editors. Advances in Neural Information Processing Systems. Vol. 22. 2009. p. 2223-2231.

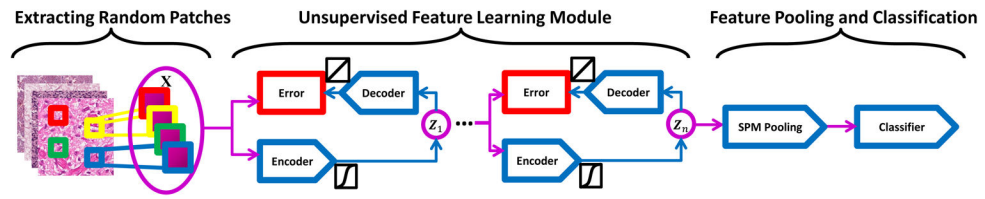


Figure 1.
Computational workflow of our approach (PSD^{SPM}).



Figure 2.
Basis functions (**B**) computed from the GBM dataset.

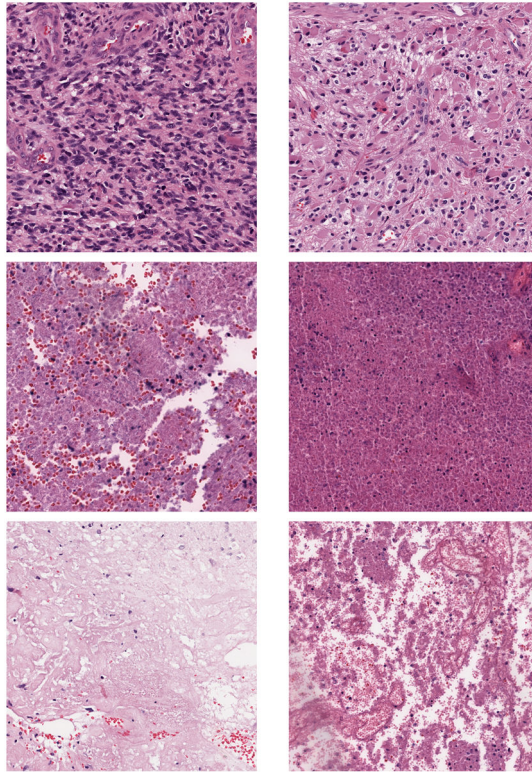


Figure 3.

GBM Examples. First row: Tumor; Second row: Transition to necrosis; Third row: Necrosis.

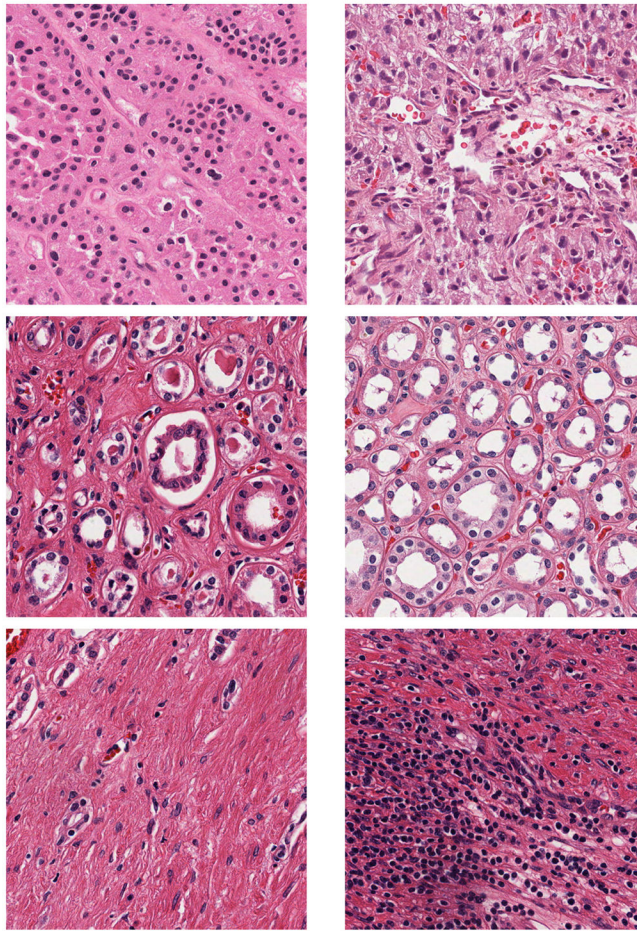


Figure 4.

KIRC Examples. First row: Tumor; Second row: Normal; Third row: Stromal.

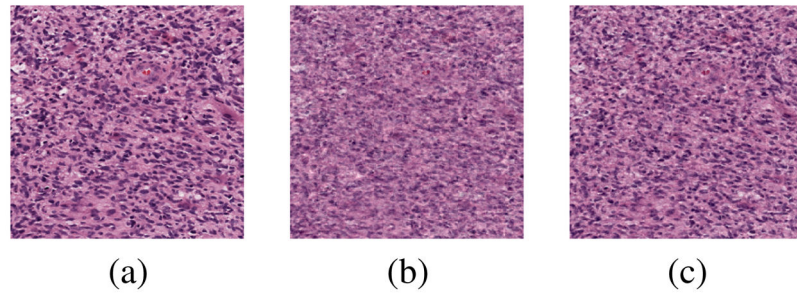


Figure 5.

Comparison of PSD with linear and nonlinear regressor 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).

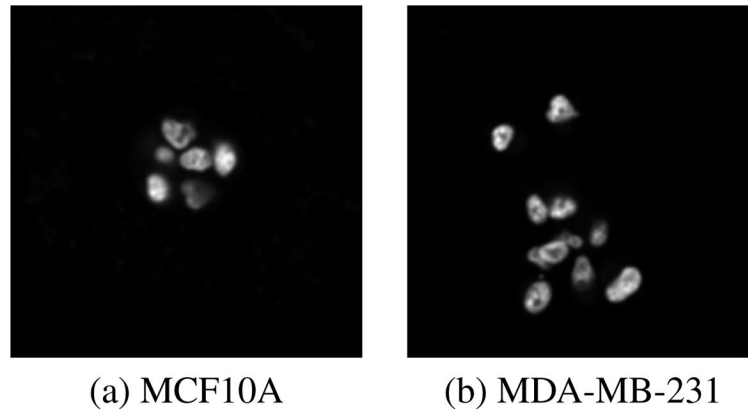


Figure 6.

Example of middle slices from two different 3D cell cultures on day 5.

Table 1

Performance of different methods on the GBM dataset.

	Method	Dictionary Size=256	Dictionary Size=512	Dictionary Size=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} [9]	91.02 ± 1.89	91.41 ± 0.95	91.20 ± 1.29
	ScSPM [34]	79.58 ± 0.61	81.29 ± 0.86	82.36 ± 1.10
	KSPM [24]	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} [9]	88.63 ± 0.91	88.91 ± 1.18	88.64 ± 1.08
	ScSPM [34]	77.65 ± 1.43	78.31 ± 1.13	81.00 ± 0.98
	KSPM [24]	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} [9]	84.06 ± 1.16	83.72 ± 1.46	83.40 ± 1.14
	ScSPM [34]	73.60 ± 1.68	75.58 ± 1.29	76.24 ± 3.05
	KSPM [24]	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 2

Performance of different methods on the KIRC dataset.

	Method	Dictionary Size=256	Dictionary Size=512	Dictionary Size=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} [9]	97.19 ± 0.49	97.27 ± 0.44	97.08 ± 0.45
	ScSPM [34]	94.52 ± 0.44	96.37 ± 0.45	96.81 ± 0.50
	KSPM [24]	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} [9]	96.80 ± 0.75	96.52 ± 0.76	96.55 ± 0.84
	ScSPM [34]	93.46 ± 0.55	95.68 ± 0.36	96.76 ± 0.63
	KSPM [24]	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} [9]	95.12 ± 0.54	95.13 ± 0.51	95.09 ± 0.40
	ScSPM [34]	91.93 ± 1.00	93.67 ± 0.72	94.86 ± 0.86
	KSPM [24]	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 3

Performance of proposed method on 3D cell culture differentiation.

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