Spectral Embedding Based Probabilistic Boosting Tree (ScEPTre): Classifying High Dimensional Heterogeneous Biomedical Data*

Pallavi Tiwari[1], Mark Rosen[2], Galen Reed[3], John Kurhanewicz[3], and Anant Madabhushi[1]

[1] Department of Biomedical Engineering, Rutgers University, USA
pallavit@eden.rutgers.edu, anantm@rci.rutgers.edu
[2] Department of Radiology, University of Pennsylvania, USA
[3] Department of Radiology, University of California, San Francisco, USA

Abstract. The major challenge with classifying high dimensional biomedical data is in identifying the appropriate feature representation to (a) overcome the curse of dimensionality, and (b) facilitate separation between the data classes. Another challenge is to integrate information from two disparate modalities, possibly existing in different dimensional spaces, for improved classification. In this paper, we present a novel data representation, integration and classification scheme, Spectral Embedding based Probabilistic boosting Tree (ScEPTre), which incorporates Spectral Embedding (SE) for data representation and integration and a Probabilistic Boosting Tree classifier for data classification. SE provides an alternate representation of the data by non-linearly transforming high dimensional data into a low dimensional embedding space such that the relative adjacencies between objects are preserved. We demonstrate the utility of ScEPTre to classify and integrate Magnetic Resonance (MR) Spectroscopy (MRS) and Imaging (MRI) data for prostate cancer detection. Area under the receiver operating Curve (AUC) obtained via randomized cross validation on 15 prostate MRI-MRS studies suggests that (a) ScEPTre on MRS significantly outperforms a Haar wavelets based classifier, (b) integration of MRI-MRS via ScEPTre performs significantly better compared to using MRI and MRS alone, and (c) data integration via ScEPTre yields superior classification results compared to combining decisions from individual classifiers (or modalities).

1 Introduction

Biomedical data such as gene expression, dynamic contrast enhanced Magnetic resonance Imaging (MRI) and Spectroscopy (MRS) suffer from the curse of dimensionality owing to the presence of large redundant, non-discriminative features within a relatively small sample space. A major challenge is to identify

* Work made possible via grants from Coulter Foundation (WHCF 4-29368), New Jersey Commission on Cancer Research, National Cancer Institute (R21CA127186-01, R03CA128081-01), and the Society for Imaging Informatics in Medicine (SIIM).

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part II, LNCS 5762, pp. 844–851, 2009.

appropriate feature representation methods and thus facilitate accurate classification. The default data representation choice for spectral and signal classification has been wavelets [1] or Principal Component Analysis (PCA) which is a linear dimensionality reduction method.

It has been shown previously that PCA fails to capture the inherent non-linear structure of high dimensional biomedical data [2]. For such cases, non linear DR methods such as Spectral Embedding (SE) [3] are more appropriate, since they reduce data dimensionality by assuming a non-linear relationship between high dimensional data samples. The objective behind SE is to non-linearly map objects $c, d \in C$ that are adjacent in the M dimensional ambient space $(\boldsymbol{F}(c), \boldsymbol{F}(d))$ to adjacent locations in the low dimensional embedding $(\boldsymbol{S}(c), \boldsymbol{S}(d))$, where $\boldsymbol{S}(c), \boldsymbol{S}(d)$ represent the β-dimensional dominant Eigen vectors corresponding to c, d $(\beta << M)$. Such non-linear mapping captures the inherent structure of the high dimensional data manifold such that object proximity and local geometries are preserved in the reduced Eigen space. These methods can also employ a wide range of similarity kernels and hence can be applied to representing diverse data including imaging, spectral, and omics.

With the wide array of multi-scale, multi-functional, multi-modal data now available for disease characterization, one of the challenges in integrated disease diagnostics is to homogeneously represent the different data streams (eg. imaging, spectroscopy, omics) to enable data fusion and classification. For example, consider the difficulties in fusing T2-weighted (w) MRI data (structural information) with MRS data (metabolic information) which involves combining scalar MR image intensity information with a high dimensional metabolic vector from the same spatial location. Data-fusion algorithms are classified broadly as data level integration and decision level integration [4]. At data level integration, original features $(\boldsymbol{F}_A(c))$ and $(\boldsymbol{F}_B(c))$ from two disparate modalities A and B are combined either via vector concatenation $\boldsymbol{F}_{AB}(c) = [\boldsymbol{F}_A(c), \boldsymbol{F}_B(c)]$ or some sort of averaging. In decision level integration, individual classifications $(\mathbf{h}_A)$ and $(\mathbf{h}_B)$ from each modality A, B are combined. Decision level integration strategies however tend to implicitly treat the data channels as independent and may result in sub optimal fusion and classification.

2 Novel Contributions of This Work

In this paper, we present an integrated data representation, fusion and classification scheme, **S**pectral **E**mbedding based **P**robabilisitic Boosting **T**ree (ScEPTre), that enables (a) homogeneous representation of multiple, disparate, high dimensional biomedical data modalities in a reduced dimensional space, and (b) fusion of disparate features from heterogeneous modalities of differing dimensionalities to obtain improved classification. Data representation and fusion in ScEPTre is performed using Spectral Embedding (SE), while the Probabilistic Boosting Tree (PBT) algorithm is used for classification. The PBT classifier generates a tree structure by recursively training each node of the tree using AdaBoost such that each node is a strong classifier [5]. PBT has the advantage of computing discriminative probabilities where hard decisions cannot be made.

In this work, we demonstrate the applicability of ScEPTre to first automatically represent and classify prostate MRS and MRI individually, followed by the data fusion of the two modalities for prostate cancer (CaP) detection. We also demonstrate that ScEPTre on MRS for CaP detection significantly outperforms classification via wavelets, the traditional method of data representation for spectral data. The contributions of this work are:

- A novel data representation scheme (ScEPTre) which enables data fusion and classification by first homogeneously representing heterogeneous data with different dimensionality within the spectrally embedded space.
- ScEPTre is shown to perform better in terms of representing high dimensional spectral data compared to the wavelet representation and yields better classification compared to decision level integration.

3 Methodological Description of ScEPTre

3.1 Spectral Embedding of High Dimensional Biomedical Data

The aim of Spectral Embedding [3] is to find an embedding vector $\boldsymbol{S}^{SE}(c_i)$, $\forall c_i \in C$, $i \in \{1, \ldots, |C|\}$, such that the relative ordering of the distances between objects in high dimensional space is maximally preserved in the lower dimensional space. Thus, if locations $c_i, c_j \in C, i, j \in \{1, \ldots, |C|\}$, are adjacent in the high dimensional feature space $\boldsymbol{F}(c_i), \boldsymbol{F}(c_j)$ respectively, then $||\boldsymbol{S}^{SE}(c_i) - \boldsymbol{S}^{SE}(c_j)||_2$ should be small, where $||.||_2$ represents the Euclidean norm. This will only be true if the distances between all $c_i, c_j \in C$ are preserved in the low dimensional mapping of the data. To compute the optimal embedding, we first define adjacency matrix $W_{SE} \in \Re^{|C|\times|C|}$ as

$$W_{SE}(i,j) = e^{-||\boldsymbol{F}(c_i)-\boldsymbol{F}(c_j)||_2}, \forall c_i, c_j \in C, i, j \in \{1, \ldots, |C|\}\,. \tag{1}$$

$\boldsymbol{S}^{SE}(c_i)$ is then obtained from the maximization of the function:

$$E(\mathcal{X}_{SE}) = 2\gamma \times trace\left[\frac{\mathcal{X}_{SE}(D - W_{SE})\mathcal{X}_{SE}^{\mathsf{T}}}{\mathcal{X}_{SE} D \mathcal{X}_{SE}^{\mathsf{T}}}\right], \tag{2}$$

where $\mathcal{X}_{SE} = \left[\boldsymbol{S}^{SE}(c_1); \boldsymbol{S}^{SE}(c_2); \ldots; \mathbf{S}^{SE}(c_n)\right], n = |C|$ and $\gamma = |C| - 1$. Additionally, D is a diagonal matrix where $\forall c \in C$, the diagonal element is defined as $D(i,i) = \sum_j W_{SE}(i,j)$. The embedding space is defined by the Eigenvectors corresponding to the smallest m Eigenvalues of $(D - W_{SE})\ \mathcal{X}_{SE} = \lambda D \mathcal{X}_{SE}$. The matrix $\mathcal{X}_{SE} \in \Re^{|C|\times\beta}$ of the first β Eigenvectors is constructed, and $\forall c_i \in C$, $\boldsymbol{S}^{SE}(c_i)$ is defined as row i of $\mathcal{X}_{SE}$.

3.2 Multi-modal Data Fusion in the Embedding Space

In [4], it has been suggested that data level integration can be achieved by aggregating features from two disparate sources ($\boldsymbol{F}_1(c)$ and $\boldsymbol{F}_2(c)$) into a single

feature vector $\boldsymbol{E}(c)$ before classification. While this may be a reasonable strategy when $|\boldsymbol{F}_1(c)|=|\boldsymbol{F}_2(c)|$, it may not be optimal when the feature vectors are of very different dimensionalities. ScEPTre enables data fusion between different dimensional feature vectors $\boldsymbol{F}_1(c)$, $\boldsymbol{F}_2(c)$ by first independently spectrally embedding them ($|\boldsymbol{S}_1^{SE}(c)| = |\boldsymbol{S}_2^{SE}(c)|$) so that the combined feature vector $[\boldsymbol{S}_1(c), \boldsymbol{S}_2(c)]$ can then be used for classification.

3.3 PBT Classifier

The PBT algorithm [5] is a combination of the Adaboost and decision trees classifiers. It iteratively generates a tree structure of length L in the training stage where each node of the tree is boosted with T weak classifiers. The hierarchical tree is obtained by dividing new samples in two subsets of $\tilde{\boldsymbol{F}}_{Right}$ and $\tilde{\boldsymbol{F}}_{Left}$ and recursively training the left and right sub-trees using Adaboost. To solve for overfitting, error parameter ϵ is introduced such that samples falling in the range $[\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon]$ are assigned to both subtrees with probabilities $(\boldsymbol{F}(c), p(1|c)) \rightarrow \tilde{\boldsymbol{F}}_{Right}(c)$ and $(\boldsymbol{F}(c), p(0|c)) \rightarrow \tilde{\boldsymbol{F}}_{Left}(c)$, where the function $p(Y|c)$ represents the posterior class conditional probability of c belonging to class $Y \in \{0, 1\}$. The algorithm stops when misclassification error hits a pre-defined threshold θ.

During testing, the conditional probability of the sample is calculated at each node based on the learned hierarchical tree. The discriminative model is obtained at the top of the tree by combining the probabilities associated with probability propagation of the sample at various nodes yielding a posterior conditional probability value $p(1|c)$, $p(0|c) \in [0, 1]$, for each sample c as belonging to one of the two classes.

4 ScEPTre for Prostate Cancer Detection Using Integrated MR Imaging and Spectroscopy

4.1 Notation and Data Description

A total of 15 1.5 Tesla (T) T2-w MRI and corresponding MRS studies were obtained prior to radical prostatectomy from University of California, San Francisco. We represent the 3D prostate T2-w scene by $\hat{\mathcal{C}} = (\hat{C}, \hat{f})$, where $\hat{C}$ is a 3D grid of voxels $\hat{c} \in \hat{C}$ and $\hat{f}(\hat{c})$ is a function that assigns an intensity value to every $\hat{c} \in \hat{C}$. We also define a spectral scene $\mathcal{C} = (C, \boldsymbol{F})$ where C is a 3D grid of MRS metavoxels, $c \in C$ and $\boldsymbol{F}$ is a spectral vector associated with each $c \in C$. Note that multiple voxels are present within the region R_{cd} between any two adjacent metavoxels $c, d \in C$. For the sake of convenience we represent R_{cd} as $R(c)$, where $\hat{c} \in R(c)$. Figure 1(a) shows a MRS spectral grid superposed on a T2-w MRI slice with expert annotated class labels $Y(c) \in \{1, 2, 3, 4, 5\}$ based on a clinical standardized 5-point scale [6] which classifies each spectra as definitely benign (1), probably benign (2), equivocal (3), probably cancer (4) and definitely cancer (5). In this work, all spectra labeled (4, 5) were assumed to be CaP and all spectra labeled as (1, 2) were assumed as benign. The 15 studies comprised 1331 class 1, 2 and 407 class 4, 5 spectra.

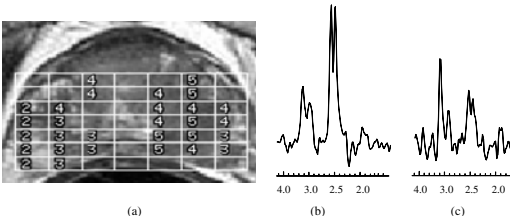


Fig. 1. (a) MRS metavoxels $c \in C$ superposed on the corresponding T2-w MRI section. Representative spectra from (b) class 2 (probably benign) and (c) class 4 (probably CaP) are shown. Note that classifying such similar looking spectra manually is a challenging, error prone, and laborious task for radiologists.

4.2 Feature Extraction of MRI and MRS

(a) Feature Extraction from MRS. Instead of extracting features derived from MRS spectral peaks, we consider the spectra in its totality. Thus, each $c \in C$, $\boldsymbol{F}(c) = [f_\alpha(c)|\alpha \in \{1, ...U\}]$ represents the MR spectral vector, reflecting the frequency component of each of the U metabolites.

(b) Feature Extraction from MRI. 38 texture features scenes were extracted motivated by previous demonstration of their utility in discriminating between the CaP and non-CaP classes [7]. We calculated the feature scenes $\hat{\mathcal{G}}_u = (\hat{C}, \hat{f}_u)$ for each $\hat{\mathcal{C}}$ by applying the feature operators $\Phi_u, u \in \{1, \ldots, 38\}$ within a local neighborhood associated with every $\hat{c} \in \hat{C}$. Hence $\hat{f}_u(\hat{c})$ is the feature value associated with feature operator Φ_u at voxel $\hat{c}$. 13 gradient, 12 first order statistical and 13 Haralick features were extracted at each $\hat{c} \in \hat{C}$. We define a T2-w MRI texture feature vector for each metavoxel $c \in C$ by taking the average of the feature values within the corresponding metavoxel as $g_u(c) = \frac{1}{|R(c)|} \sum_{\hat{c} \in R(c)} \left[\hat{f}_u(\hat{c})\right]$ where $|R(c)|$ represents the cardinality of the set of voxels contained in the space between any 2 adjacent meta-voxels. The corresponding feature vector is then given as $\boldsymbol{G}(c) = [g_u(c)|u \in \{1, \ldots, 38\}], \forall c \in C$.

4.3 Data Representation, Fusion, and Classification via ScEPTre

(a) Data Representation. For each $c \in C$, $\boldsymbol{G}(c)$, $\boldsymbol{F}(c)$ is reduced to β dimensional feature vectors $\boldsymbol{S}^{T2}$ and $\boldsymbol{S}^{MRS}$, corresponding to the spectrally embedded T2-w and MRS vectors [8].

(b) Data Fusion. Owing to the physical and dimensional differences in the MRS and T2-w MRI features, the MRS-MRI meta-classifier is created in the joint T2-w MRI and MRS embedding space where the physicality of the object features has been removed. A direct concatenation of the T2-w MRI and MRS embedding coordinates can be obtained as $\boldsymbol{S}^{T2MRS} = [\boldsymbol{S}^{T2}(c), \boldsymbol{S}^{MRS}(c)]$. The concatenated feature vector $\boldsymbol{S}^{T2MRS}(c)$ is then used for classification.

(c) Classification. PBT generates a posterior conditional probability for CaP, $p(w_T|\ \boldsymbol{S}^{\phi})$, $\phi \in \{T2, MRS, T2MRS\}$ for each spectral location $c \in C$, based on the embedding vector $\boldsymbol{S}^{\phi}(c)$, where w_T represents CaP class (4, 5 spectra). We define $\mathbf{h}^{\phi,\rho}(c)$ as the binary prediction result at each threshold $\rho \in [0, 1]$ such that $\mathbf{h}^{\phi,\rho}(c) = 1$ when $p(w_T|\boldsymbol{S}^{\phi}) \geq \rho$, 0 otherwise.

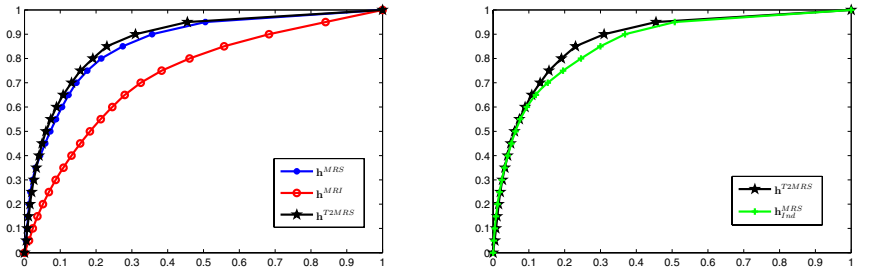
PBT was trained using $\theta = 0.45$ and $\epsilon = 0.4$ as suggested in [5]. Five weak classifiers were used to train each node of PBT using AdaBoost for length $L = 5$. Since samples for each class (1, 2 and 4, 5) were not equally distributed within the dataset, the number of training samples ($\Pi \in \{100, 200, 300\}$) from each class was varied to evaluate the classifier with respect to training and the remaining were used for testing. A randomized cross validation strategy comprising 25 runs was used to evaluate PBT performance for each set of training data ($\Pi \in \{100, 200, 300\}$). For each training set, Receiver Operating Characteristic (ROC) curves were computed and mean μ^{AUC} and standard deviation σ^{AUC} of area under curve (AUC) were computed over 25 runs. The experiments were repeated for different numbers of embedding dimensions ($\beta \in \{5, 10, 15\}$).

5 Results and Discussion

Figure 2(a) shows average ROC curves obtained for $\mathbf{h}^{T2}$, $\mathbf{h}^{MRS}$, and $\mathbf{h}^{T2MRS}$ across 25 runs of randomized cross validation for $\beta = 10$ using 300 training samples from each class (1, 2 and 4, 5). The highest AUC value corresponds to the classifier $\mathbf{h}^{T2MRS}$(shown in black), while the lowest is for $\mathbf{h}^{T2}$ (shown in red). Figure 2(b) shows average ROC curves obtained from $\mathbf{h}^{T2MRS}$ and $\hat{\mathbf{h}}^{T2MRS}$ for $\beta = 10$ using 300 training samples from each class across 25 cross validation runs. AUC and accuracy values for $\mathbf{h}^{T2}$, $\mathbf{h}^{MRS}$, and $\mathbf{h}^{T2MRS}$ averaged over 25 cross validation runs are summarized in Table 1 with corresponding standard deviations for $\beta = \{5, 10, 15\}$. Our quantitative evaluation demonstrate that CaP classification obtained via ScEPTre employing multi-modal integration of MRI-MRS outperforms classification obtained via, (a) wavelets, (b) MRI, MRS alone, and (c) decision level integration of MRI-MRS.

5.1 MRS via ScEPTre and Wavelet Based PBT Classification

Discrete wavelet transform (DWT) based representation of MRS was compared against the spectral embedding representation. DWT transforms a M dimensional spectral signal $\boldsymbol{F}(c)$ into a feature vector $\boldsymbol{S}^{WT}(c)$ containing M wavelet coefficients using a set of basis functions. Each wavelet coefficient is calculated by taking the dot product of spectral vector $\boldsymbol{F}(c)$ with one of the N basis functions, derived from the mother wavelet by a series of translations and dilations. The top m of M largest wavelet coefficients were used for classification with the PBT to distinguish between cancer (4, 5) and benign spectra (1, 2). Different values of $N \in \{3, 4, 5, 6, 7, 8, 9\}$ were employed in the spectral representation. Quantitative results indicate that ScEPTre on MRS performs better compared to a wavelet-based PBT classifier, both in terms of accuracy and AUC. ScEPTre using $\mathbf{h}^{MRS}$ results in μ^{AUC}=0.8743 ($\beta = 10$, $\Pi = 300$) compared to $\mathbf{h}^{WT}$ ($N = 7$, $\Pi = 300$) which yields $\mu^{AUC} = 0.7725$.



(a) (b)

Fig. 2. Average ROC curves across 300 training samples over 25 cross validation runs for classifiers (a) $\mathbf{h}^{T2}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{T2MRS}$, and (b) $\mathbf{h}^{T2MRS}$, $\hat{\mathbf{h}}^{T2MRS}$. The best performance in both (a) and (b) corresponds to the classifier (shown in black) based on data level integration of structural and metabolic data ($\mathbf{h}^{T2MRS}$).

Table 1. Table shows the μ^{AUC}, σ^{AUC} and accuracy results for 300 training samples from each class using ScEPTre on $\mathbf{h}^{T2}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{T2MRS}$ respectively, for $\beta \in \{5, 10, 15\}$

β	Accuracy			AUC		
	$\mathbf{h}^{T2}$	$\mathbf{h}^{MRS}$	$\mathbf{h}^{T2MRS}$	$\mathbf{h}^{T2}$	$\mathbf{h}^{MRS}$	$\mathbf{h}^{T2MRS}$
5	56.53 ± 5.53	65.35 ± 4.5	67.62 ± 3.5	65.13 ± 3.13	76.32 ± 1.87	75.82 ± 2.09
10	67.72±3.42	79.44±3.33	**79.51±2.54**	73.27±2.36	86.17±1.91	**87.43±1.41**
15	71.12± 4.47	81.52 ± 2.53	**83.64 ± 2.61**	73.89 ± 1.73	87.49 ± 1.49	**90.27 ± 1.52**

5.2 ScEPTre Based MRI-MRS Meta-classification Compared to MRI/MRS Alone

As is apparent from Figure 2 and Table 1, MRI-MRS data fusion ($\mathbf{h}^{T2MRS}$) using ScEPTre significantly outperformed classification obtained by both T2-w MRI ($\mathbf{h}^{T2}$) and MRS ($\mathbf{h}^{MRS}$) individually.

5.3 Data Level Integration vs. Decision Level Integration

The higher AUC values for $\mathbf{h}^{T2MRS}$ compared to $\hat{\mathbf{h}}^{T2MRS}$ suggests that data level integration is superior to decision level classification. Paired student t-tests were also conducted for AUC at the operating point of the average ROC curves, with the null hypothesis being no improvement in performance of $\mathbf{h}^{T2MRS}$ when compared to the other 3 classifiers ($\mathbf{h}^{T2}$, $\mathbf{h}^{MRS}$, $\hat{\mathbf{h}}^{T2MRS}$). Significantly superior performance ($p < 0.05$) was observed for $\mathbf{h}^{T2MRS}$ suggesting that integrating structural textural features and metabolic information at the data-level offers the most optimal results for CaP detection.

6 Concluding Remarks and Future Directions

In this paper, we presented ScEPTre, an integrated data representation, integration and classification scheme which is capable of (a) accurately representing

high dimensional biomedical data, (b) integrating two heterogeneous multimodal datasets embedded in different dimensionalities, and (c) yielding superior data-level integration compared to decision level combination. Data representation was employed using spectral embedding, a non-linear dimensionality reduction scheme which extracts meaningful class relationships embedded in the data. Data integration in ScEPTre is then performed by concatenating Eigen feature vectors obtained from the different modalities, subsequently classified via a Probabilistic Boosting Tree. In this work we presented the application of ScEPTre for representation, integration and classification of MRS and MRI for prostate cancer detection. Results show that ScEPTre on MRS significantly outperforms a classifier that uses wavelets to represent the MRS data. We also demonstrated that integration of MRI-MRS using ScEPTre performs better than both MRI and MRS alone, and ScEPTre based data fusion is superior to decision level integration from a classification perspective.

References

1. Unser, M., Aldroubi, A.: A review of wavelets in biomedical applications. Proceedings of the IEEE 84, 626–638 (1996)
2. Lee, G., Madabhushi, A., Rodriguez, C.: Investigating the efficacy of nonlinear dimensionality reduction schemes in classifying gene and protein expression studies. IEEE Trans. on Comp. Biol. and Bioinf. 5, 368–384 (2008)
3. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE PAMI 22(8), 888–905 (2000)
4. Rohlfing, T., Pfefferbaum, A., Sullivan, E.V., Maurer Jr., C.R.: Information fusion in biomedical image analysis: Combination of data vs. Combination of interpretations. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 150–161. Springer, Heidelberg (2005)
5. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. ICCV 2, 1589–1596 (2005)
6. Jung, J., Coakley, F., Vigneron, D., Swanson, M., Qayyum, A., Kurhanewicz, J., et al.: Prostate depiction at endorectal MR Spectroscopic Imaging: investigation of a standardized evaluation system. Radiology 233, 701–708 (2004)
7. Madabhushi, A., Feldman, M., Metexas, D., Tomaszeweski, J., Chute, D.: Automated Detection of Prostatic Adenocarcinoma from High-Resolution Ex Vivo MRI. IEEE TMI 2412, 1611–1625 (2005)
8. Tiwari, P., Madabhushi, A., Rosen, M.A.: A hierarchical unsupervised spectral clustering scheme for detection of prostate cancer from magnetic resonance spectroscopy (MRS). In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 278–286. Springer, Heidelberg (2007)