Simultaneous Sparsity Model for Histopathological Image Classification

Umamahesh Srinivas[1] Hojjat Mousavi[1] Charles Jeon[2]
Vishal Monga[1] Arthur Hattel[3] Bhushan Jayarao[3]

[1]Dept. of Electrical Engineering, Pennsylvania State University

[2]Dept. of Electrical and Systems Engineering, University of Pennsylvania

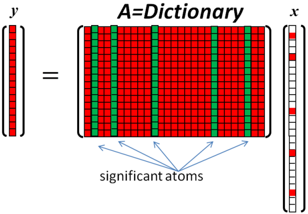
[3]Dept. of Veterinary and Biomedical Sciences, Pennsylvania State University

iPAL @ PENNSTATE
Information Processing and Algorithms Laboratory

IEEE International Symposium on Biomedical Imaging

April 10, 2013

The compressed sensing (CS) problem[2]

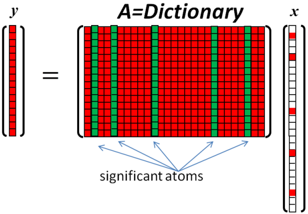


How to represent $\boldsymbol{y}$ with the fewest number of dictionary atoms?

[1] Donoho, Communications on Pure and Applied Mathematics, 2006

[2] Candes et al., IEEE Transactions on Information Theory, 2006

The compressed sensing (CS) problem[2]



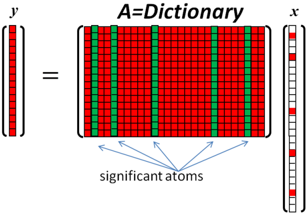
How to represent $\boldsymbol{y}$ with the fewest number of dictionary atoms?

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \text{ subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

[1] Donoho, Communications on Pure and Applied Mathematics, 2006

[2] Candes et al., IEEE Transactions on Information Theory, 2006

The compressed sensing (CS) problem[2]



How to represent $\boldsymbol{y}$ with the fewest number of dictionary atoms?

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \text{ subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

Convex relaxation[1]:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \text{ subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

[1] Donoho, Communications on Pure and Applied Mathematics, 2006

[2] Candes et al., IEEE Transactions on Information Theory, 2006

Sparse representation-based image classification[3]

- Extension of CS analytical framework to classification by designing class-specific dictionaries

[3] Wright et al., IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009

Sparse representation-based image classification[3]

- Extension of CS analytical framework to classification by designing class-specific dictionaries
- Given: M classes, with N_i training images in the i-th class
- Assumption: New image from class i lies (approximately) in the linear span of training samples from class i

$$\boldsymbol{y} = x_{i,1}\boldsymbol{a}_{i,1} + x_{i,2}\boldsymbol{a}_{i,2} + \ldots + x_{i,N_i}\boldsymbol{a}_{i,N_i} = \boldsymbol{A}_i\boldsymbol{x}_i$$

[3] Wright et al., IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009

Sparse representation-based image classification[3]

- Extension of CS analytical framework to classification by designing class-specific dictionaries
- Given: M classes, with N_i training images in the i-th class
- Assumption: New image from class i lies (approximately) in the linear span of training samples from class i

$$\boldsymbol{y} = x_{i,1}\boldsymbol{a}_{i,1} + x_{i,2}\boldsymbol{a}_{i,2} + \ldots + x_{i,N_i}\boldsymbol{a}_{i,N_i} = \boldsymbol{A}_i\boldsymbol{x}_i$$

- $\boldsymbol{y} \rightarrow$ sparse linear combination of all training samples:

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \ldots & \boldsymbol{A}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_M \end{bmatrix} = \boldsymbol{A}\boldsymbol{x}$$

[3] Wright et al., IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009

Sparse representation-based image classification[3]

- Extension of CS analytical framework to classification by designing class-specific dictionaries
- Given: M classes, with N_i training images in the i-th class
- Assumption: New image from class i lies (approximately) in the linear span of training samples from class i

$$\boldsymbol{y} = x_{i,1}\boldsymbol{a}_{i,1} + x_{i,2}\boldsymbol{a}_{i,2} + \ldots + x_{i,N_i}\boldsymbol{a}_{i,N_i} = \boldsymbol{A}_i\boldsymbol{x}_i$$

- $\boldsymbol{y} \rightarrow$ sparse linear combination of all training samples:

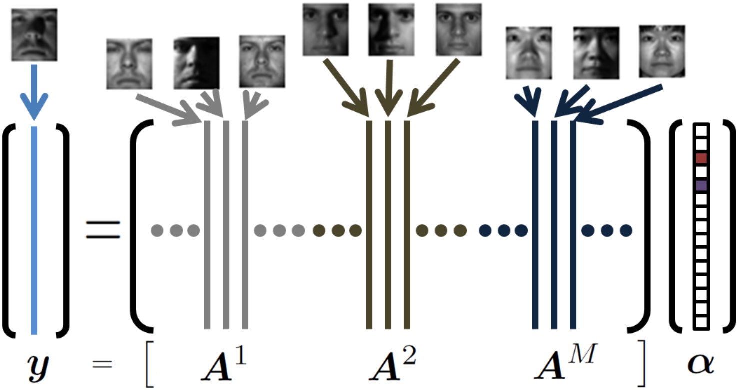
$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \ldots & \boldsymbol{A}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_M \end{bmatrix} = \boldsymbol{A}\boldsymbol{x}$$

- Membership of $\boldsymbol{y}$ encoded by sparse representation

$$\boldsymbol{x} = \begin{bmatrix} \mathbf{0}^t & \ldots & \mathbf{0}^t & \boldsymbol{x}_i^t & \mathbf{0}^t & \ldots & \mathbf{0}^t \end{bmatrix}^t.$$

[3] Wright et al., IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009

Application: Face recognition



Sparse representation-based classification (SRC)

- Solve the sparse recovery problem:

$$\hat{\boldsymbol{x}} = \arg\min \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \epsilon$$

Sparse representation-based classification (SRC)

- Solve the sparse recovery problem:

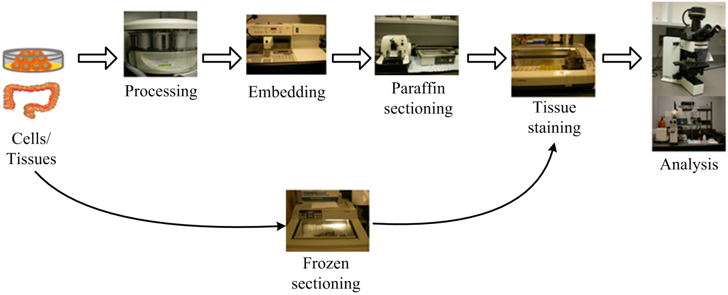
$$\hat{\boldsymbol{x}} = \arg\min \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \epsilon$$

- Class decision based on reconstruction residuals:

$$\text{identity}(\boldsymbol{y}) = \arg\min_i \|\boldsymbol{y} - \boldsymbol{A}\delta_i(\hat{\boldsymbol{x}})\|_2$$

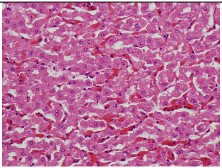
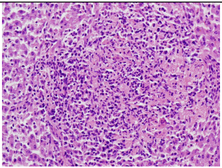
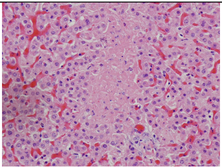
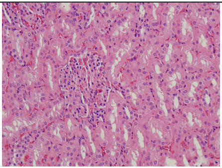
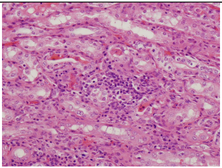
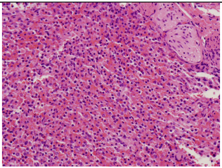
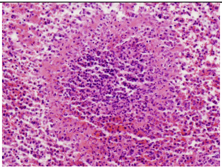
$\delta_i(\hat{\boldsymbol{x}}) \rightarrow$ only non-zero entries are those associated with class i

Histopathology



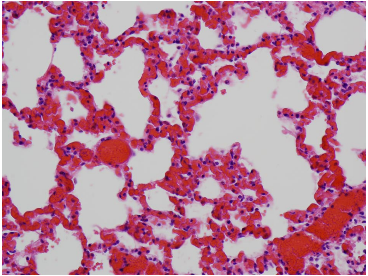
- High resolution images → underlying tissue architecture preserved
- H&E-staining
  - Hematoxylin → bluish nuclei
  - Eosin → red cytoplasm and connective tissue
- Applications: cancer or more generally disease detection and grading → breast, prostate, lung, liver, spleen, etc.

Example histopathological images[4]

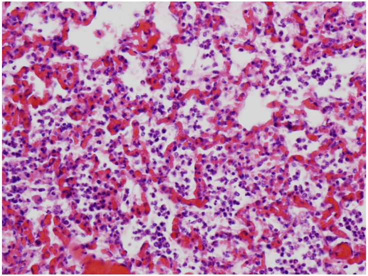
Organ	Healthy	Inflammation	Necrosis
Liver			
Kidney			-
Spleen			-

[4] Animal Diagnostics Laboratory, Penn State

Example images: Lung



(a) Healthy lung.



(b) Lung inflammation.

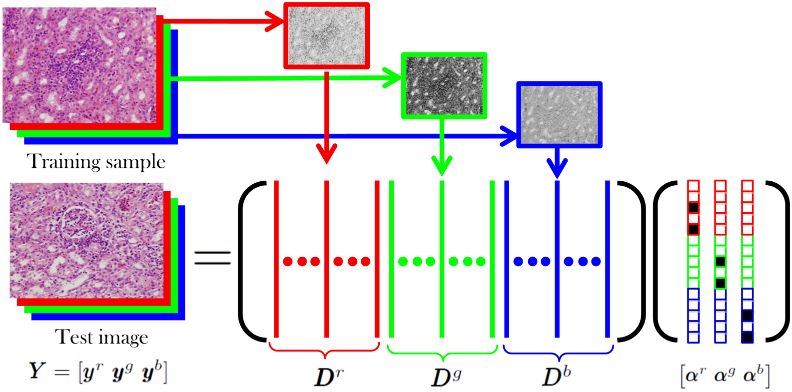
Opportunities and challenges

1. Multi-channel (color) information
    - Extension of sparsity-based techniques via color dictionaries
    - Maintain color channel correspondence

Opportunities and challenges

1. Multi-channel (color) information
   - Extension of sparsity-based techniques via color dictionaries
   - Maintain color channel correspondence
2. Rich structural information in tissues
   - Feature extraction is a challenging task → customized to specific organs
   - Example-based learning

SHIRC: Simultaneous Sparsity model for Histopathological Image Representation and Classification



Designing color channel constraints

$$\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{S} = \begin{bmatrix} \boldsymbol{D}_h^r & \boldsymbol{D}_d^r & \boldsymbol{D}_h^g & \boldsymbol{D}_d^g & \boldsymbol{D}_h^b & \boldsymbol{D}_d^b \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}^r & \boldsymbol{\alpha}^g & \boldsymbol{\alpha}^b \end{bmatrix}$$

Designing color channel constraints

$$\boldsymbol{Y} = \boldsymbol{DS} = \left[\boldsymbol{D}_h^r \;\; \boldsymbol{D}_d^r \;\; \boldsymbol{D}_h^g \;\; \boldsymbol{D}_d^g \;\; \boldsymbol{D}_h^b \;\; \boldsymbol{D}_d^b\right]\left[\boldsymbol{\alpha}^r \;\; \boldsymbol{\alpha}^g \;\; \boldsymbol{\alpha}^b\right]$$

$$\boldsymbol{\alpha}^r = \begin{bmatrix} \boldsymbol{\alpha}_h^r \\ \boldsymbol{\alpha}_d^r \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \boldsymbol{\alpha}^g = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\alpha}_h^g \\ \boldsymbol{\alpha}_d^g \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \boldsymbol{\alpha}_b = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\alpha}_h^b \\ \boldsymbol{\alpha}_d^b \end{bmatrix}$$

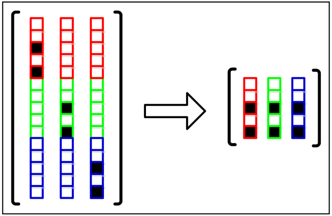
- $\boldsymbol{\alpha}^r, \boldsymbol{\alpha}^g, \boldsymbol{\alpha}^b \in \mathbb{R}^{3n}$

Designing color channel constraints

$$\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{S} = \left[\boldsymbol{D}_h^r \quad \boldsymbol{D}_d^r \quad \boldsymbol{D}_h^g \quad \boldsymbol{D}_d^g \quad \boldsymbol{D}_h^b \quad \boldsymbol{D}_d^b\right]\left[\boldsymbol{\alpha}^r \quad \boldsymbol{\alpha}^g \quad \boldsymbol{\alpha}^b\right]$$

$$\boldsymbol{\alpha}^r = \begin{bmatrix} \boldsymbol{\alpha}_h^r \\ \boldsymbol{\alpha}_d^r \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \boldsymbol{\alpha}^g = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\alpha}_h^g \\ \boldsymbol{\alpha}_d^g \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \boldsymbol{\alpha}_b = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\alpha}_h^b \\ \boldsymbol{\alpha}_d^b \end{bmatrix}$$

- $\boldsymbol{\alpha}^r, \boldsymbol{\alpha}^g, \boldsymbol{\alpha}^b \in \mathbb{R}^{3n}$

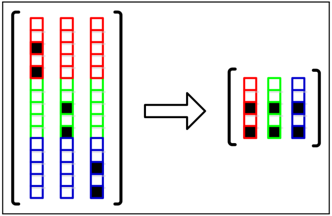


Designing color channel constraints

$$\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{S} = \begin{bmatrix} \boldsymbol{D}_h^r & \boldsymbol{D}_d^r & \boldsymbol{D}_h^g & \boldsymbol{D}_d^g & \boldsymbol{D}_h^b & \boldsymbol{D}_d^b \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}^r & \boldsymbol{\alpha}^g & \boldsymbol{\alpha}^b \end{bmatrix}$$

$$\boldsymbol{\alpha}^r = \begin{bmatrix} \boldsymbol{\alpha}_h^r \\ \boldsymbol{\alpha}_d^r \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \boldsymbol{\alpha}^g = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\alpha}_h^g \\ \boldsymbol{\alpha}_d^g \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \boldsymbol{\alpha}_b = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\alpha}_h^b \\ \boldsymbol{\alpha}_d^b \end{bmatrix}$$

- $\boldsymbol{\alpha}^r, \boldsymbol{\alpha}^g, \boldsymbol{\alpha}^b \in \mathbb{R}^{3n}$



$$\boldsymbol{S}' = \begin{bmatrix} \boldsymbol{\alpha}_h^r & \boldsymbol{\alpha}_h^g & \boldsymbol{\alpha}_h^b \\ \boldsymbol{\alpha}_d^r & \boldsymbol{\alpha}_d^g & \boldsymbol{\alpha}_d^b \end{bmatrix}$$

$$\boldsymbol{S} \in \mathbb{R}^{3n \times 3} \xrightarrow{?} \boldsymbol{S}' \in \mathbb{R}^{n \times 3}$$

Formulating the optimization problem

Define:

$$\boldsymbol{H} = \begin{bmatrix} \mathbf{1}_n & \mathbf{0}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{1}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{1}_n \end{bmatrix} \in \mathbb{R}^{3n \times 3}, \; \boldsymbol{J} = [\boldsymbol{I}_n \quad \boldsymbol{I}_n \quad \boldsymbol{I}_n] \in \mathbb{R}^{n \times 3n}$$

Formulating the optimization problem

Define:

$$\boldsymbol{H} = \begin{bmatrix} \mathbf{1}_n & \mathbf{0}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{1}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{1}_n \end{bmatrix} \in \mathbb{R}^{3n \times 3}, \; \boldsymbol{J} = [\boldsymbol{I}_n \quad \boldsymbol{I}_n \quad \boldsymbol{I}_n] \in \mathbb{R}^{n \times 3n}$$

Hadamard operator:

$$\boldsymbol{H} \circ \boldsymbol{S} = \begin{bmatrix} \boldsymbol{\alpha}^r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\alpha}^g & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}^b \end{bmatrix} \in \mathbb{R}^{3n \times 3}$$

Formulating the optimization problem

Define:

$$\boldsymbol{H} = \begin{bmatrix} \mathbf{1}_n & \mathbf{0}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{1}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{1}_n \end{bmatrix} \in \mathbb{R}^{3n \times 3}, \ \boldsymbol{J} = [\boldsymbol{I}_n \quad \boldsymbol{I}_n \quad \boldsymbol{I}_n] \in \mathbb{R}^{n \times 3n}$$

Hadamard operator:

$$\boldsymbol{H} \circ \boldsymbol{S} = \begin{bmatrix} \boldsymbol{\alpha}^r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\alpha}^g & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}^b \end{bmatrix} \in \mathbb{R}^{3n \times 3}$$

$$\boldsymbol{J}\left(\boldsymbol{H} \circ \boldsymbol{S}\right) = [\boldsymbol{I}_n \quad \boldsymbol{I}_n \quad \boldsymbol{I}_n] \begin{bmatrix} \boldsymbol{\alpha}^r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\alpha}^g & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}^b \end{bmatrix} = \left[\boldsymbol{\alpha}^r \quad \boldsymbol{\alpha}^g \quad \boldsymbol{\alpha}^b\right] = \boldsymbol{S}'$$

Formulating the optimization problem

Define:

$$\boldsymbol{H} = \begin{bmatrix} \mathbf{1}_n & \mathbf{0}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{1}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{1}_n \end{bmatrix} \in \mathbb{R}^{3n\times 3},\ \boldsymbol{J} = [\boldsymbol{I}_n \quad \boldsymbol{I}_n \quad \boldsymbol{I}_n] \in \mathbb{R}^{n\times 3n}$$

Hadamard operator:

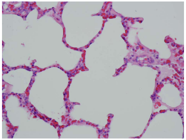
$$\boldsymbol{H} \circ \boldsymbol{S} = \begin{bmatrix} \boldsymbol{\alpha}^r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\alpha}^g & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}^b \end{bmatrix} \in \mathbb{R}^{3n\times 3}$$

$$\boldsymbol{J}\left(\boldsymbol{H} \circ \boldsymbol{S}\right) = [\boldsymbol{I}_n \quad \boldsymbol{I}_n \quad \boldsymbol{I}_n] \begin{bmatrix} \boldsymbol{\alpha}^r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\alpha}^g & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}^b \end{bmatrix} = \left[\boldsymbol{\alpha}^r \quad \boldsymbol{\alpha}^g \quad \boldsymbol{\alpha}^b\right] = \boldsymbol{S}'$$

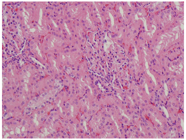
$$\hat{\boldsymbol{S}}' = \arg\min \|\boldsymbol{S}'\|_{\text{row},0} \text{ subject to } \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{S}\|_F \leq \epsilon$$

Experimental set-up: Data sets

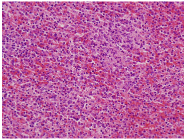
- ADL data set[5]
  - Mammalian tissue
  - Three organs: kidney, lung, spleen (healthy vs. inflammation)
  - 40 training images, 80 test images



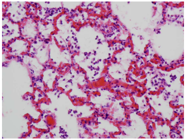
(a) Lung



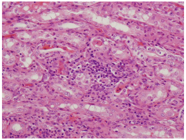
(b) Kidney



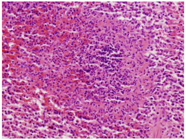
(c) Spleen



(d) Lung



(e) Kidney



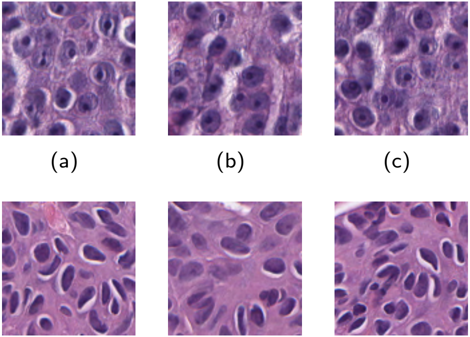
(f) Spleen

Figure : Top row: healthy; bottom row: inflammatory.

5 Animal Diagnostics Laboratory, Penn State

Experimental set-up: Data sets

- IBL data set[6]
    - Human intraductal breast lesions
    - Ductal carcinoma in situ (DCIS) vs. usual ductal hyperplasia (UDH)
    - 32 training images, 22 test images



(a) (b) (c)

(d) (e) (f)

Figure : Top row: DCIS; bottom row: UDH.

[6] Clarian Pathology Lab, Indianapolis; Computer and Information Science Dept., IUPUI

Experimental set-up

Comparison with:

1. SRC: using luminance channel information only
2. SVM: State-of-the-art WND-CHARM features[7,8]
   - Image pixel statistics
   - Color channel histograms
   - Gray level co-occurrence matrix (GLCM)
   - Wavelet coefficients
   - Morphological features

[7] Orlov et al., Pattern Recognition Letters, 2008

[8] http://ome.grc.nia.nig.gov/wnd-charm

Results I: Confusion matrix

- Each row: true class identity of image
- Each column: class predicted by the classifier

Table : Lung (ADL)

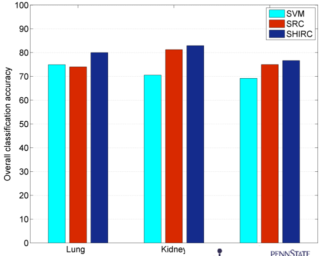
Class	Healthy	Inflammatory	Method
Healthy	**0.8875**	0.1125	SVM
	0.7250	0.2750	SRC
	0.7500	0.2500	SHIRC
Inflammatory	0.3762	0.6238	SVM
	0.2417	0.7583	SRC
	0.1500	**0.8500**	SHIRC

Table : Kidney (ADL)

Class	Healthy	Inflammatory	Method
Healthy	0.6925	0.3075	SVM
	0.8750	0.1250	SRC
	0.8250	0.1750	SHIRC
Inflammatory	0.2812	0.7188	SVM
	0.2500	0.7500	SRC
	0.1667	**0.8333**	SHIRC

Table : Spleen (ADL)

Class	Healthy	Inflammatory	Method
Healthy	0.5112	0.4888	SVM
	0.7083	0.2917	SRC
	0.6500	0.3500	SHIRC
Inflammatory	0.1275	0.8725	SVM
	0.2083	0.7917	SRC
	0.1167	**0.8833**	SHIRC



Results I: Confusion matrix

- Each row: true class identity of image
- Each column: class predicted by the classifier

Table : Lung (ADL)

Class	Healthy	Inflammatory	Method
Healthy	**0.8875**	0.1125	SVM
	0.7250	0.2750	SRC
	0.7500	0.2500	SHIRC
Inflammatory	0.3762	0.6238	SVM
	0.2417	0.7583	SRC
	0.1500	**0.8500**	SHIRC

Table : Kidney (ADL)

Class	Healthy	Inflammatory	Method
Healthy	0.6925	0.3075	SVM
	0.8750	0.1250	SRC
	0.8250	0.1750	SHIRC
Inflammatory	0.2812	0.7188	SVM
	0.2500	0.7500	SRC
	0.1667	**0.8333**	SHIRC

Table : Spleen (ADL)

Class	Healthy	Inflammatory	Method
Healthy	0.5112	0.4888	SVM
	0.7083	0.2917	SRC
	0.6500	0.3500	SHIRC
Inflammatory	0.1275	0.8725	SVM
	0.2083	0.7917	SRC
	0.1167	**0.8833**	SHIRC

Table : IBL[a]

Class	UDH	DCIS	Method
UDH	0.7140	0.2860	SRC
	0.8540	0.1460	SHIRC
DCIS	0.1750	0.8250	SRC
	0.1220	**0.8780**	SHIRC

[a] Dundar et al., IEEE Transactions on Biomedical Engineering, 2011

Results II: ROC curves

- Miss: diseased identified as healthy
- False alarm: healthy classified as diseased

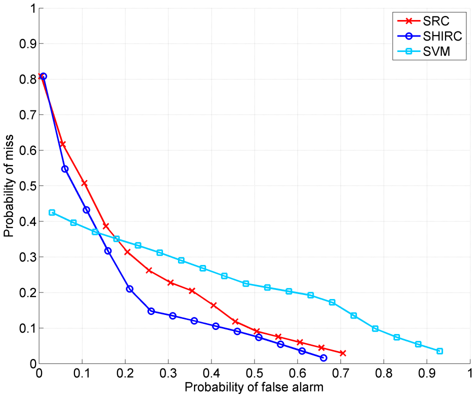


Figure : Lung (ADL).

Results II: ROC curves

- Miss: diseased identified as healthy
- False alarm: healthy classified as diseased

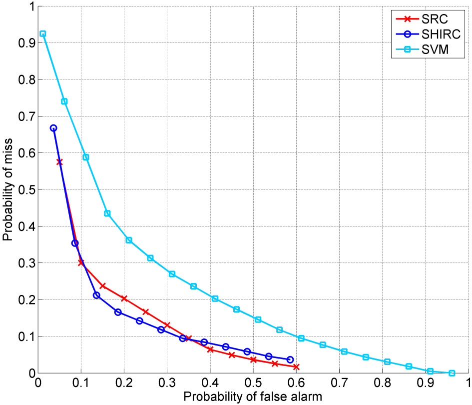


Figure : Kidney (ADL).

Results II: ROC curves

- Miss: diseased identified as healthy
- False alarm: healthy classified as diseased

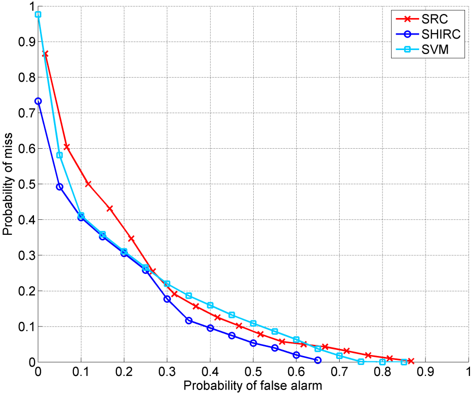


Figure : Spleen (ADL).

Conclusions

- Joint sparsity models incorporate contextual information
  - Color channel information in medical imaging
- New greedy algorithm[9] for multi-dictionary sparsity-based classification

Ongoing/Future research:

- Histopathological objects of interest (cells, nuclei) at different sizes and scales $\rightarrow$ locally adaptive framework
- Effect of training set size on classification accuracy.

[9] MATLAB code: `signal.ee.psu.edu/histimg.html`

Thank you

Questions?