RDECOM

Joint Sparsity for Target Detection

TECHNOLOGY DRIVEN. WARFIGHTER FOCUSED.

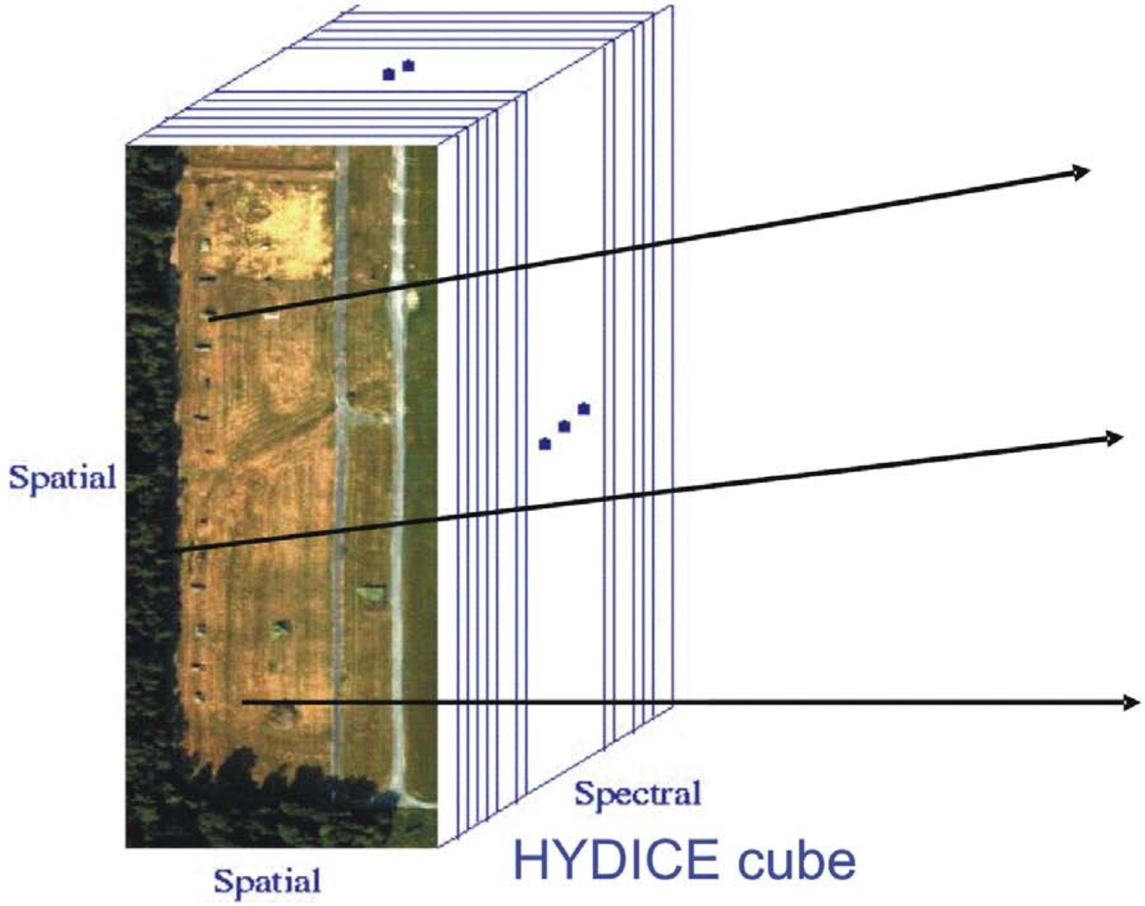
Nasser M. Nasrabadi

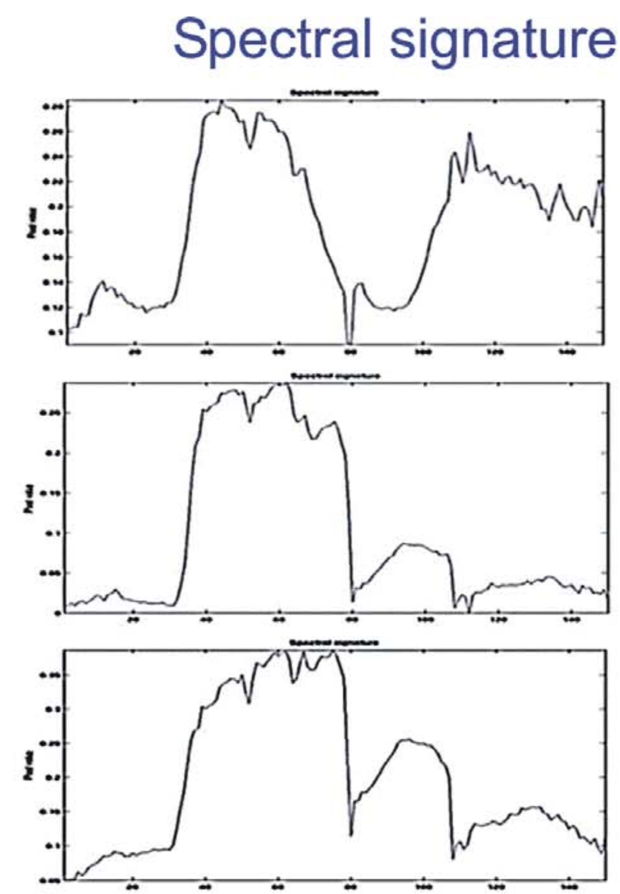
U.S. Army Research Laboratory

Introduction

- Objective: Segmentation of HSI into multiple classes (target and background) or classify individual objects (military targets) from multiple views of the same physical target.
- Assumptions
  - Training data: known spectral characteristics (or images) of different classes
  - Test data: a sparse linear combination of all training data
  - In HSI Neighboring pixels: similar materials
  - Mutiple views of targets are similar
- Results compared to classical SVM classifiers

Hyperspectral Imagery





Pixel-wise Sparsity Model

- Background pixels approximately lie in a low-dimensional subspace

$$\boldsymbol{x}_i = \alpha_{i,1}^b \boldsymbol{a}_1^b + \alpha_{i,2}^b \boldsymbol{a}_2^b + \cdots + \alpha_{i,N_b}^b \boldsymbol{a}_{N_b}^b = \boldsymbol{A}^b \boldsymbol{\alpha}_i^b$$

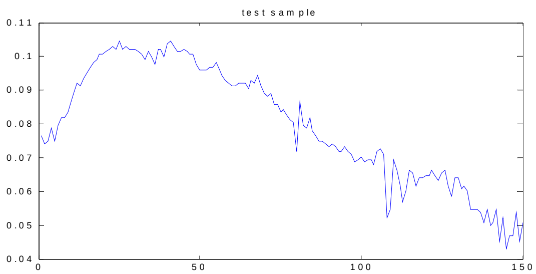
- Target pixels also lie in a low-dimensional subspace

$$\boldsymbol{x}_i = \alpha_{i,1}^t \boldsymbol{a}_1^t + \alpha_{i,2}^t \boldsymbol{a}_2^t + \cdots + \alpha_{i,N_t}^t \boldsymbol{a}_{N_t}^t = \boldsymbol{A}^t \boldsymbol{\alpha}_i^t$$

- A test sample $\boldsymbol{x}_i$ can be sparsely represented by

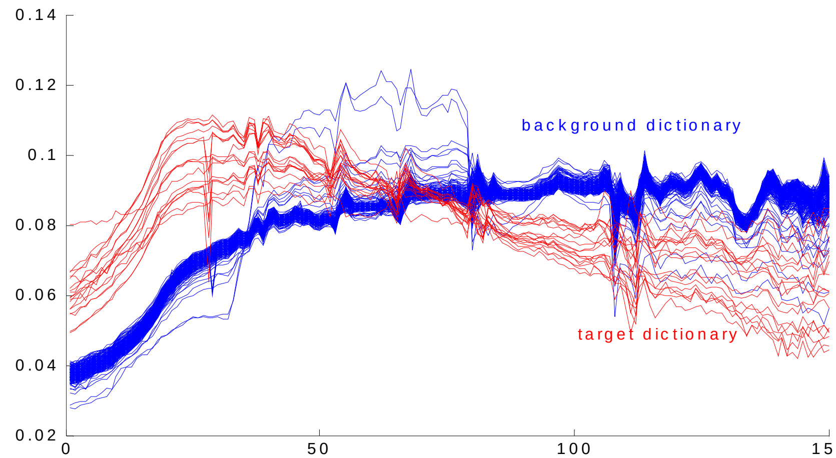
$$\boldsymbol{x}_i = \boldsymbol{A}^b \boldsymbol{\alpha}_i^b + \boldsymbol{A}^t \boldsymbol{\alpha}_i^t = \begin{bmatrix} \boldsymbol{A}^b & \boldsymbol{A}^t \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_i^b \\ \boldsymbol{\alpha}_i^t \end{bmatrix} = \boldsymbol{A} \boldsymbol{\alpha}_i$$

Illustration: Pixel-Wise Sparse Model

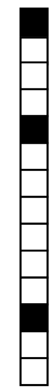


Target Pixel

=



×



Test Spectrum $\boldsymbol{x}_i$

Spectral dictionary $\boldsymbol{A}$

■ Nonzero entries

$$\boldsymbol{x}_i = \boldsymbol{A}^b \boldsymbol{\alpha}_i^b + \boldsymbol{A}^t \boldsymbol{\alpha}_i^t = \begin{bmatrix} \boldsymbol{A}^b & \boldsymbol{A}^t \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_i^b \\ \boldsymbol{\alpha}_i^t \end{bmatrix} = \boldsymbol{A}\boldsymbol{\alpha}_i$$

Sparse Recovery

- Sparse coefficient is recovered by

$$\hat{\boldsymbol{\alpha}}_i = \arg\min \|\boldsymbol{\alpha}_i\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{\alpha}_i = \boldsymbol{x}_i$$

- For empirical data

$$\hat{\boldsymbol{\alpha}}_i = \arg\min \|\boldsymbol{\alpha}_i\|_0 \quad \text{subject to} \quad \|\boldsymbol{A}\boldsymbol{\alpha}_i - \boldsymbol{x}_i\|_2 \le \sigma$$

$$\hat{\boldsymbol{\alpha}}_i = \arg\min \|\boldsymbol{A}\boldsymbol{\alpha}_i - \boldsymbol{x}_i\|_2 \quad \text{subject to} \quad \|\boldsymbol{\alpha}_i\|_0 \le K_0$$

- NP-hard problem
  - Greedy algorithms: MP, OMP, SP, CoSaMP, LARS
  - Convex relaxation: Iterative Thresholding, Primal-Dual Interior-Point, Gradient Projection, Proximal Gradient, Augmented Lagrange Multiplier

$$\hat{\boldsymbol{\alpha}}_i = \arg\min \|\boldsymbol{\alpha}_i\|_1 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{\alpha}_i = \boldsymbol{x}_i$$

Classification Based on Residuals

- Recover sparse coefficient $\hat{\boldsymbol{\alpha}}_i = \begin{bmatrix} \hat{\boldsymbol{\alpha}}_i^b \\ \hat{\boldsymbol{\alpha}}_i^t \end{bmatrix}$
- Compute the residuals (approximation errors w.r.t. the two sub-dictionaries)

$$r_b\left(\boldsymbol{x}_i\right) = \left\|\boldsymbol{x}_i - \boldsymbol{A}^b \hat{\boldsymbol{\alpha}}_i^b\right\|_2 \quad \text{and} \quad r_t\left(\boldsymbol{x}_i\right) = \left\|\boldsymbol{x}_i - \boldsymbol{A}^t \hat{\boldsymbol{\alpha}}_i^t\right\|_2$$

- Class of test pixel $\boldsymbol{x}_i$ is made by comparing the residuals

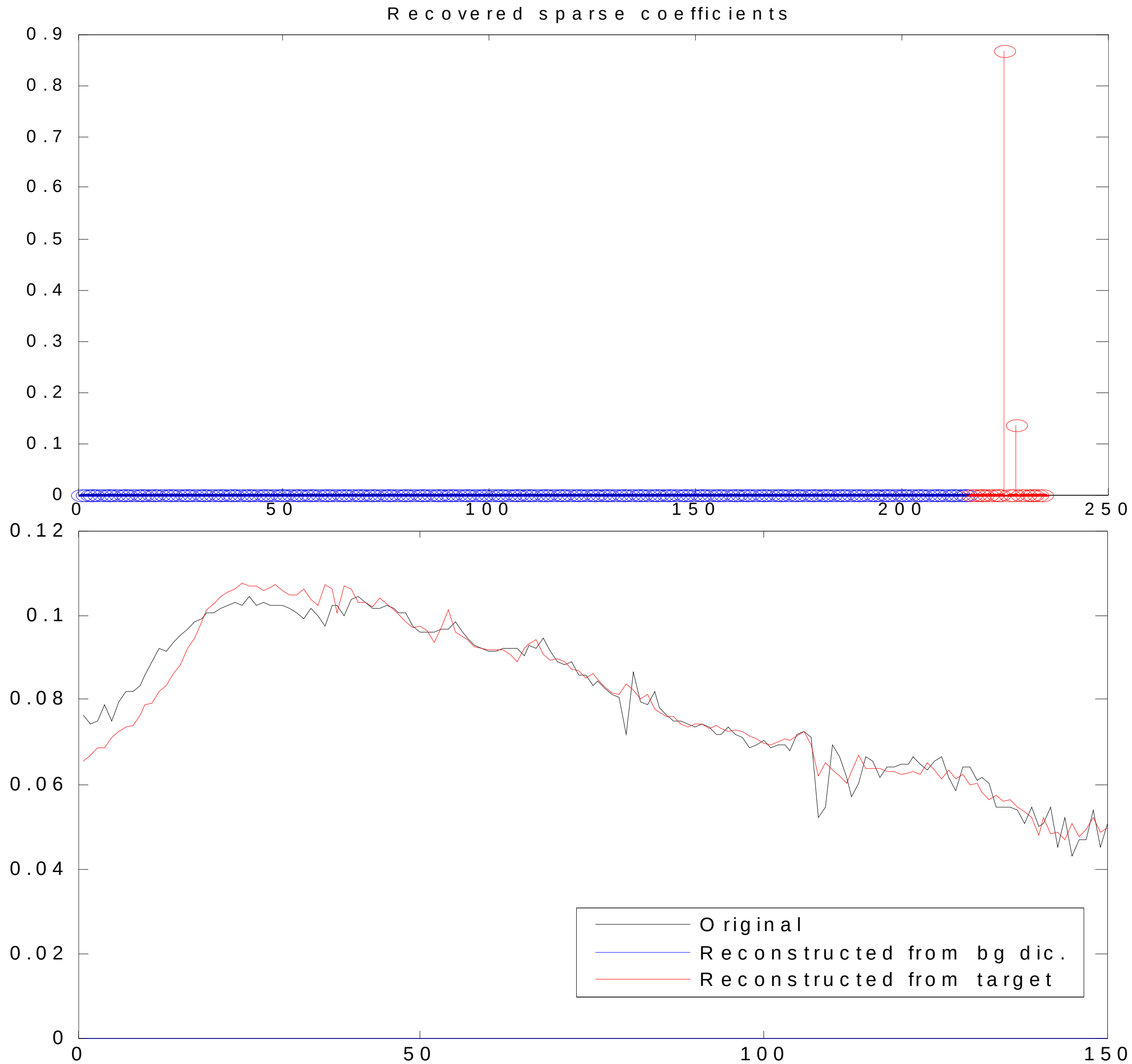
$$r_b(\boldsymbol{x}_i) - r_t(\boldsymbol{x}_i) \underset{\text{background}}{\overset{\text{target}}{\gtrless}} \delta$$

Example: Reconstruction

$$\hat{\alpha}_i = \begin{bmatrix} \hat{\alpha}_i^b \\ \hat{\alpha}_i^t \end{bmatrix}$$

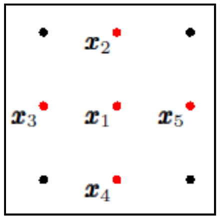
$$\boxed{\hat{x}_i^t = A^t \hat{\alpha}_i^t}$$

$$\boxed{\hat{x}_i^b = A^b \hat{\alpha}_i^b}$$



Joint Sparsity Model
(Joint Structural Sparsity Prior)

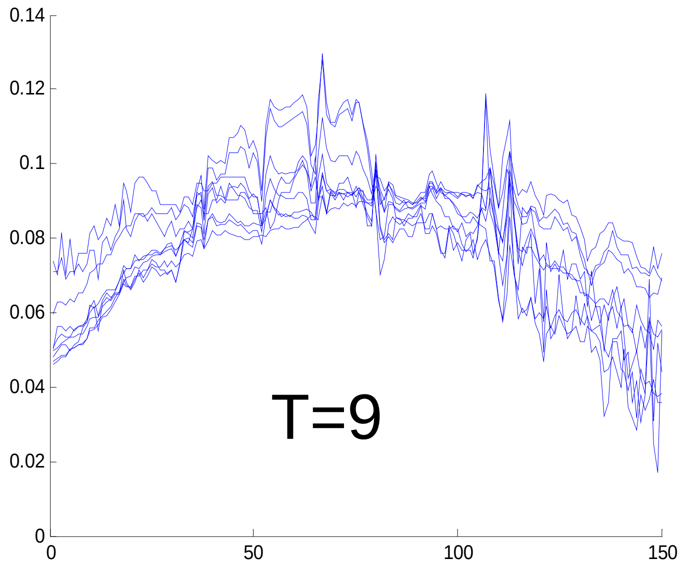
- Use of contextual information
  - Neighboring pixels: similar spectral characteristics
  - Approximated by the same few training samples, weighed differently
- Consider *T* pixels in a small neighborhood



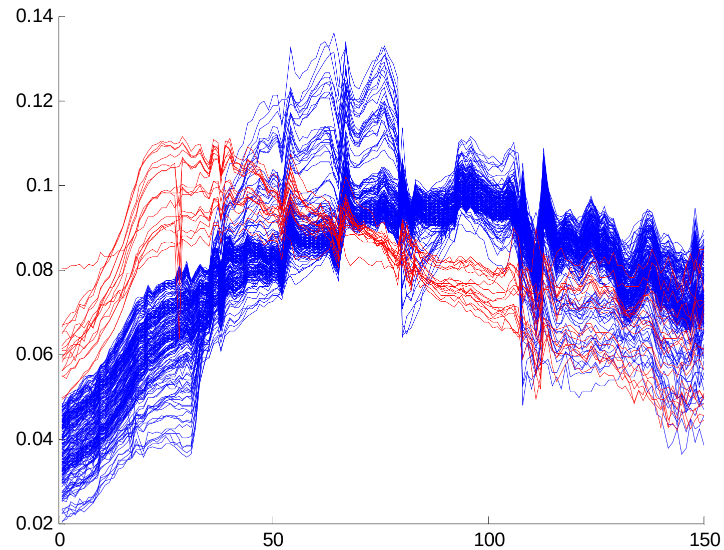
$$
\left.\begin{aligned}
\boldsymbol{x}_1 &= \boldsymbol{A}\boldsymbol{\alpha}_1 \\
\boldsymbol{x}_2 &= \boldsymbol{A}\boldsymbol{\alpha}_2 \\
&\vdots \\
\boldsymbol{x}_T &= \boldsymbol{A}\boldsymbol{\alpha}_T
\end{aligned}\right.
\Rightarrow \boldsymbol{X} = [\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \cdots \quad \boldsymbol{x}_T] = \boldsymbol{A}\underbrace{[\boldsymbol{\alpha}_1 \quad \boldsymbol{\alpha}_2 \quad \cdots \quad \boldsymbol{\alpha}_T]}_{\boldsymbol{S}} = \boldsymbol{A}\boldsymbol{S}
$$

  - $\boldsymbol{\alpha}_i$'s: sparse vectors with same support, different magnitude
  - $\boldsymbol{S}$: sparse matrix with only a few nonzero rows

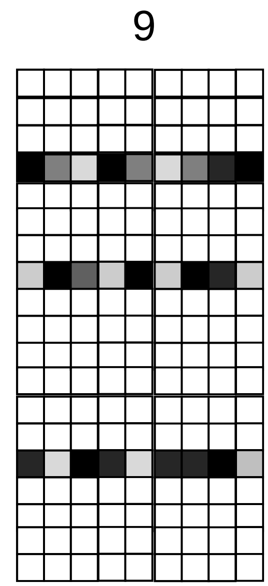
Illustration: T=3x3 Neighborhood



$=$



$\times$



Data matrix $\boldsymbol{X}$

Spectral dictionary $\boldsymbol{A}$

Row-sparse matrix $\boldsymbol{S}$

Joint Sparse Recovery

- $\boldsymbol{S}$ is recovered by

$$\hat{\boldsymbol{S}} = \arg\min \left\| \boldsymbol{S} \right\|_{\text{row, } 0} \quad \text{subject to} \quad \boldsymbol{AS} = \boldsymbol{X}$$

- Solved by greedy algorithms: Simultaneous OMP (SOMP) , Simultaneous SP (SSP) or Convex optimization to find the same active set

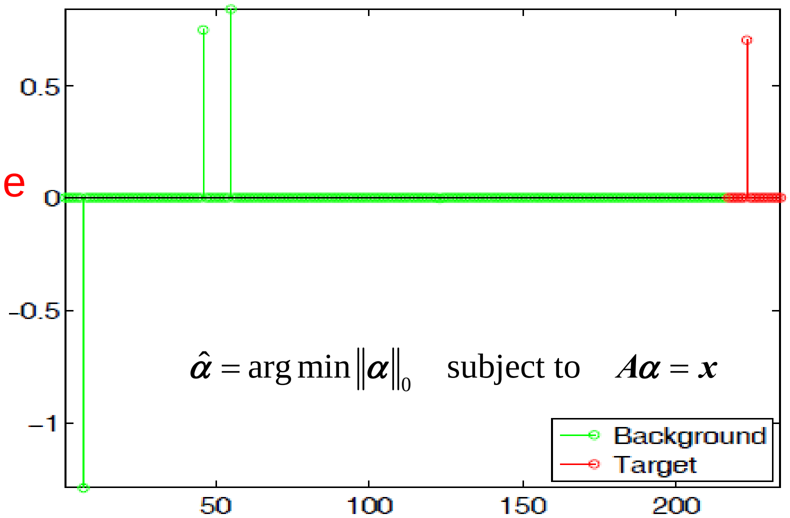
$$\hat{\boldsymbol{S}} = \arg\min \left\| \boldsymbol{S} \right\|_{1,2} \quad \text{subject to} \quad \boldsymbol{AS} = \boldsymbol{X}$$

- Decision obtained by comparing total residuals

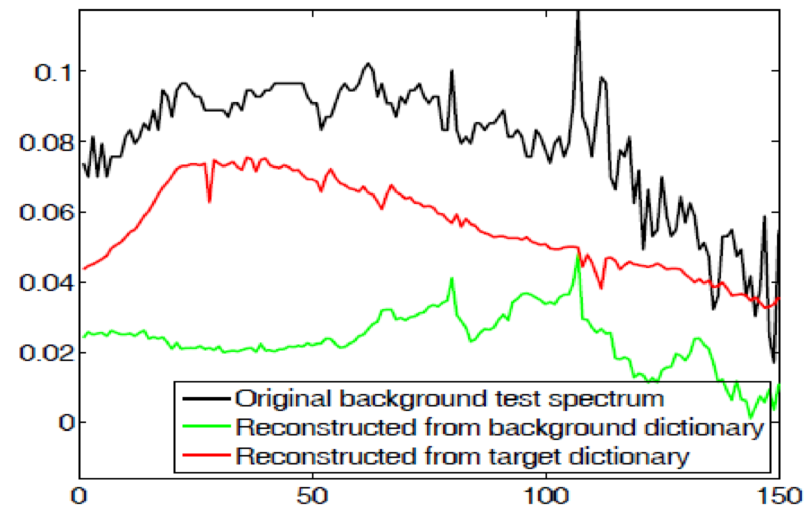
$$\left\| \boldsymbol{X} - \boldsymbol{A}^b \hat{\boldsymbol{S}}^b \right\|_F - \left\| \boldsymbol{X} - \boldsymbol{A}^t \hat{\boldsymbol{S}}^t \right\|_F \underset{\text{background}}{\overset{\text{target}}{\gtrless}} \delta$$

Comparison of single pixel sparsity model VS Joint Sparsity Recovery Model (k=5 atoms active)

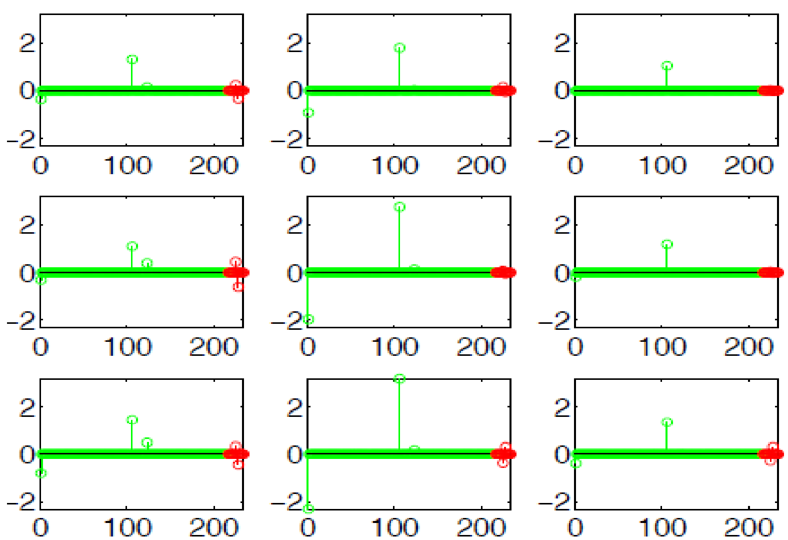
Input a single background pixel x

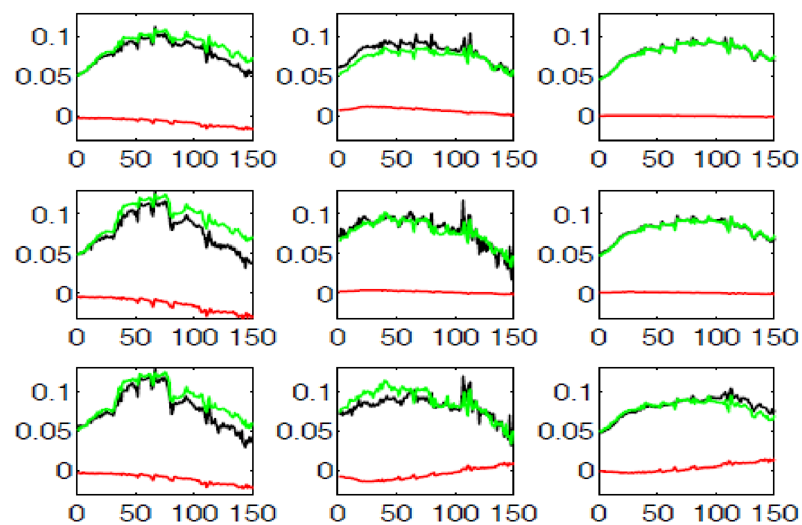


$$\hat{\alpha} = \arg\min \|\alpha\|_0 \quad \text{subject to} \quad A\alpha = x$$



Input nine neighboring background pixels X





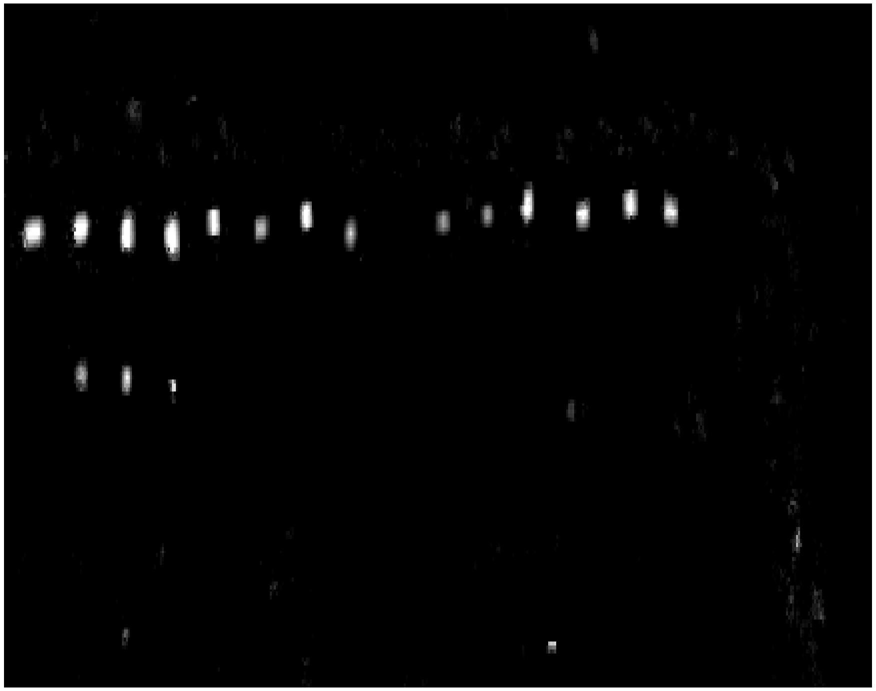
$$\hat{S} = \arg\min \|S\|_{\text{row},0} \quad \text{subject to} \quad AS = X$$

RDECOM

Results on HYDICE FR-I

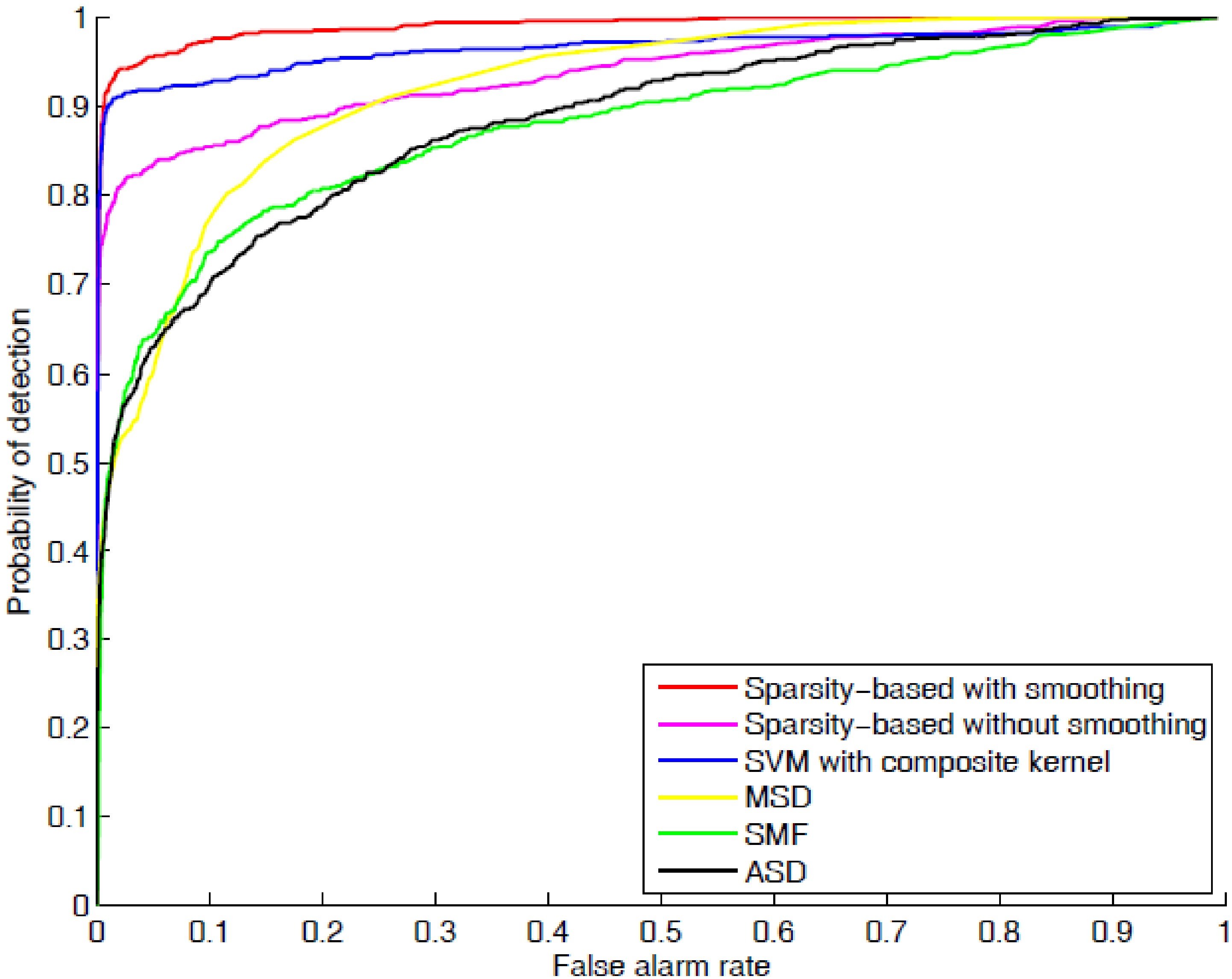
ARL

Original image (averaged over 150 bands)



Proposed detector output

Results on FR-I: ROC Curves



Extension to Multiple Classes

- AVIRIS HSI data set with 16 classes, 220 bands, 20 meters pixel resolution

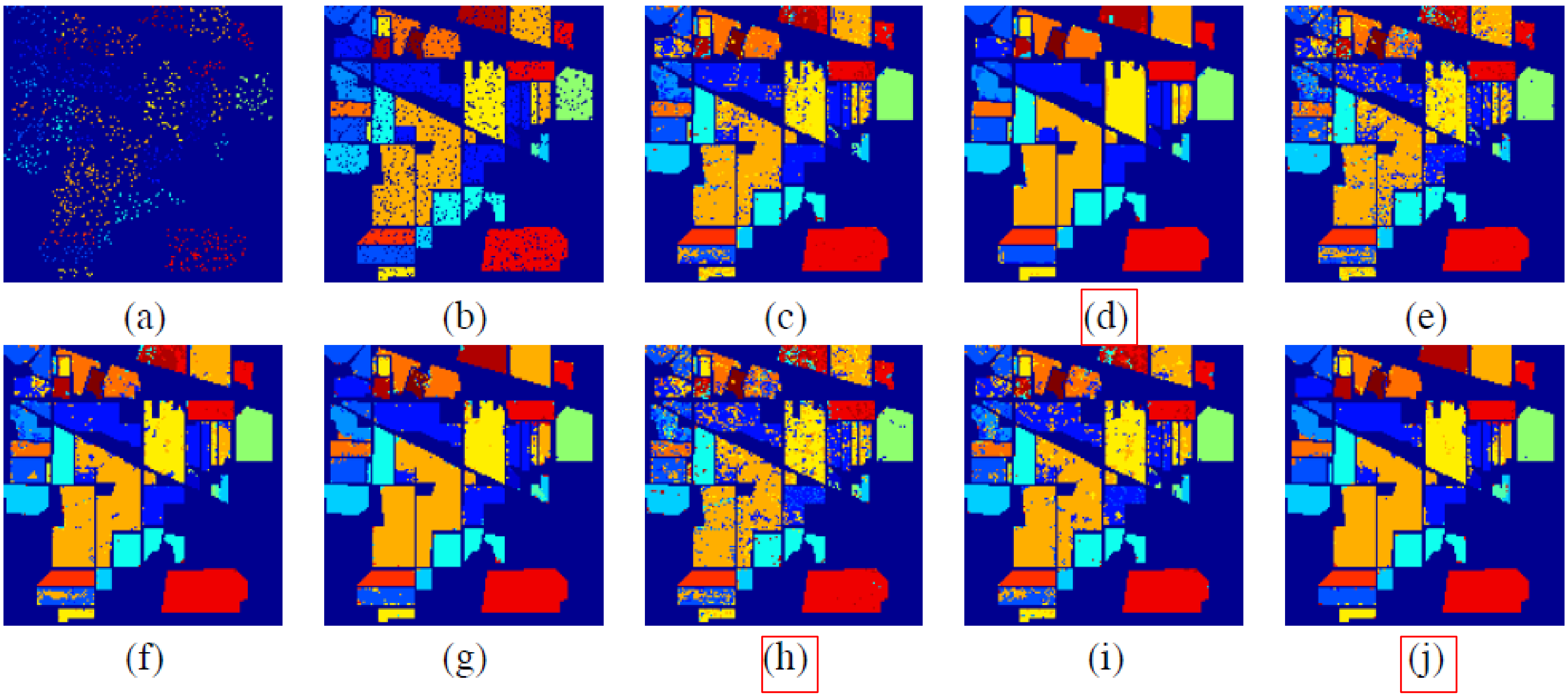


Fig. 2. For the Indian Pines image: (a) training set and (b) test set. Classification maps obtained by (c) SVM, (d) SVM-CK, (e) SP, (f) SP-S, (g) SSP, (h) OMP, (i) OMP-S, and (j) SOMP.

Extension to Multiple Classes

CLASSIFICATION ACCURACY (%) FOR THE INDIAN PINES IMAGE ON THE TEST SET.

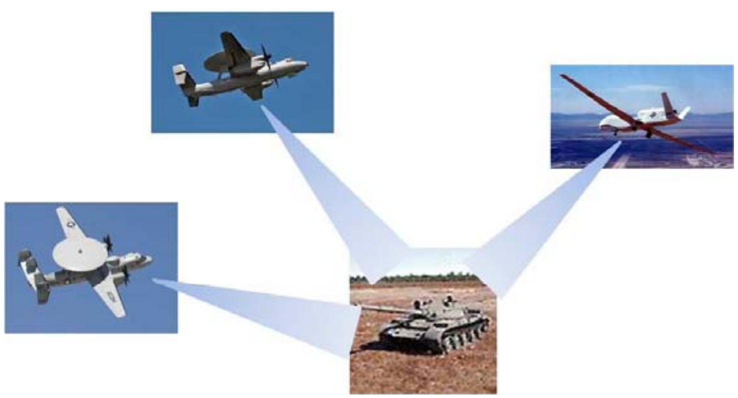
Class	SVM	SVM-CK	SP	SP-S	SSP	OMP	OMP-S	SOMP	ℓ_1
1	81.25	**95.83**	68.75	87.50	89.58	68.75	70.83	85.42	39.58
2	86.28	**96.67**	74.65	91.94	95.04	65.97	79.22	94.88	78.53
3	72.80	90.93	63.20	82.53	92.93	60.67	76.67	**94.93**	51.87
4	58.10	85.71	40.00	70.95	85.24	38.57	55.24	**91.43**	28.57
5	92.39	93.74	89.04	94.41	92.17	89.49	**95.30**	89.49	80.76
6	96.88	97.32	95.98	99.26	98.81	95.24	98.96	98.51	**99.40**
7	43.48	69.57	21.74	47.83	73.91	21.74	52.17	**91.30**	17.39
8	98.86	98.41	99.09	**99.77**	99.55	97.05	**99.77**	99.55	99.32
9	50.00	55.56	61.11	**94.44**	0	33.33	72.22	0	16.67
10	71.53	**93.80**	70.72	86.80	88.98	68.20	82.32	89.44	63.95
11	84.38	94.37	77.94	93.38	**97.34**	75.96	88.79	**97.34**	86.04
12	85.51	**93.66**	61.23	84.24	86.59	54.53	73.73	88.22	57.79
13	**100**	99.47	**100**	**100**	99.47	**100**	98.95	**100**	**100**
14	93.30	**99.14**	95.62	98.28	98.88	92.87	97.25	**99.14**	97.94
15	64.91	87.43	48.25	69.30	97.37	41.23	49.71	**99.12**	35.96
16	88.24	**100**	92.94	95.29	85.88	94.12	**100**	96.47	90.59
Overall	84.52	94.86	78.10	91.16	94.79	74.78	85.52	**95.28**	77.99
Average	79.24	**90.73**	72.52	87.25	86.36	68.61	80.70	88.45	65.27
κ	0.823	0.941	0.749	0.899	0.940	0.712	0.834	**0.946**	0.746

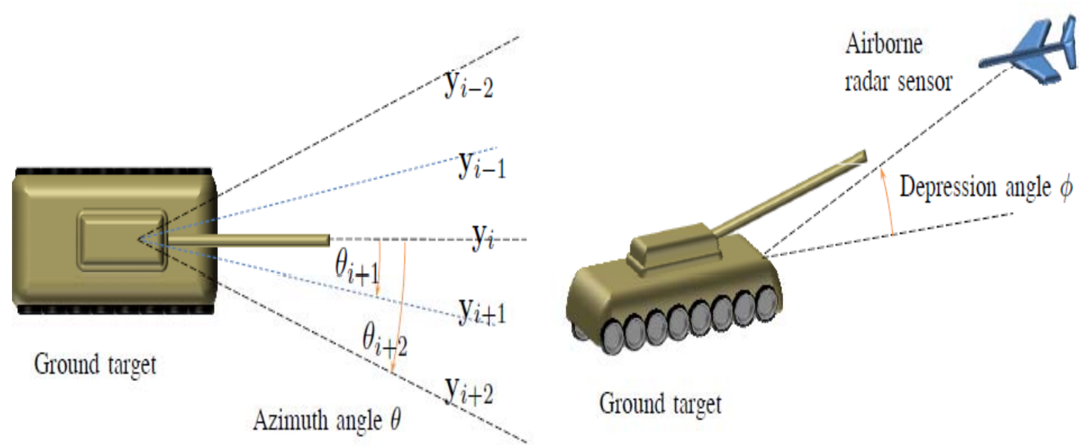
Multi-View Target Classification

- In ATR applications we can have multiple observations of the same physical target from different platforms or from the same platform at different viewing angles (aspects).

$$\hat{\boldsymbol{\alpha}}_i = \arg\min \|\boldsymbol{\alpha}_i\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{\alpha}_i = \boldsymbol{y}_i \qquad \text{(Single-Measurement)}$$

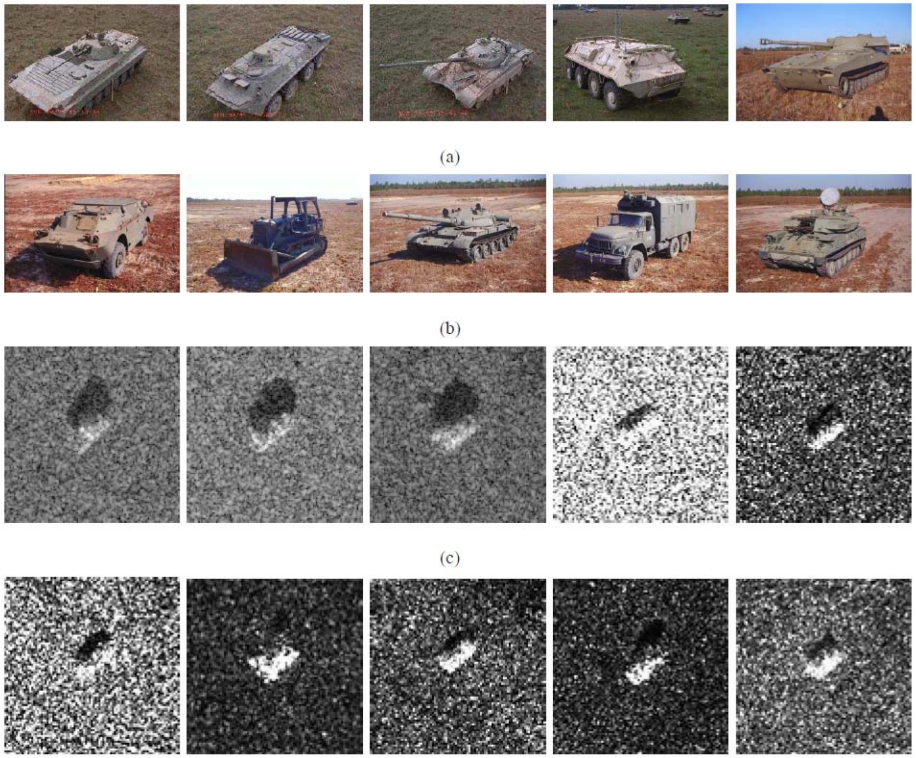
$$\hat{\boldsymbol{S}} = \arg\min \|\boldsymbol{S}\|_{\text{row, 0}} \quad \text{subject to} \quad \boldsymbol{AS} = \boldsymbol{Y} \qquad \text{(Multi-Measurements)}$$





Experimental Results on Multi-View Target Classification

- MSTAR SAR data-base consists of 10 military targets at roughly 1-3° interval azimuth angles (0-360°) at two different depression angles 15° and 17°. Data from 17° is used for training (dictionary design) 15° is used for testing



(a) (b) (c) (d)

SUMMARY OF THE MSATR DATABASE.

No.	1			2	3			4	5	6	7	8	9	10	
Type	BMP2 sn-c9563	sn-c9566	sn-c21	BTR70	T72 sn-132	sn-812	sn-S7	BTR60	2S1	BRDM2	D7	T62	ZIL131	ZSU23/4	SUM
Train (17° dep) [1]	233	[232]	[233]	233	232	[231]	[228]	256	299	298	299	299	299	299	2747
Test (15° dep)	195	196	196	196	196	195	191	195	274	274	274	273	274	274	3203

Experimental Results on Multi-View Target Classification

- Three class (BMP2, BTR70, T72) target classification C=3 with multiple views M=3 . Features are incoherent random projections dimension range from d=128 to1024.

COMPARISON OF SINGLE-VIEW AND MULTI-VIEW ATR ACCURACY (%) ON THE MSTAR DATABASE UNDER DIFFERENT FEATURE DIMENSIONS ($M = 3$, $C = 3$, $N = 698$)

Feature Dims. (d)	128	256	512	1024	Avg.
Single-view SVM	63.7	66.8	76.8	81.2	72.1
Single-view SRC	65.6	78.2	89.5	92.3	81.4
Single-view KSVM	76.2	81.1	85.4	88.9	82.9
SVM	73.2	76.6	81.9	86.9	79.6
SRC	73.8	86.3	92.9	95.5	87.1
KSVM	**83.2**	87.2	88.6	91.6	87.7
JSRC	80.3	**89.7**	**94.2**	**95.6**	**90.0**

$$\hat{\boldsymbol{\alpha}}_i = \arg\min \|\boldsymbol{\alpha}_i\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{\alpha}_i = \boldsymbol{x}_i$$

$$\hat{\boldsymbol{\beta}} = \arg\min \|\boldsymbol{\beta}\|_0 \quad \text{subject to} \quad \tilde{\boldsymbol{A}}\boldsymbol{\beta} = \tilde{\boldsymbol{x}}$$

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_M \end{bmatrix} \quad \text{and} \quad \tilde{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_M \end{bmatrix}$$

$$\hat{\boldsymbol{S}} = \arg\min \|\boldsymbol{S}\|_{\text{row},0} \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{S} = \boldsymbol{X}$$

Note $\boldsymbol{S} = [\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_M]$

Experimental Results on Number of Views and Angle Size

- Effect of different number of views *M*

MULTI-VIEW ATR ACCURACY (%) ON THE MSTAR DATABASE WITH DIFFERENT NUMBER OF VIEWS ($d = 256$, $C = 3$, $N = 698$).

Views (M)	2	3	4	5	6	7	8	9	10	11	12	13	Avg.
SVM	71.5	76.6	77.6	80.1	81.3	82.6	82.3	82.9	83.7	84.1	84.5	84.7	81.0
SRC	84.5	86.3	88.8	89.0	89.0	90.3	90.6	91.5	92.1	93.2	93.1	93.9	90.2
KSVM	85.5	87.2	88.3	89.4	90.4	90.9	91.1	92.0	91.8	92.4	92.8	93.1	90.4
JSRC	86.9	89.7	91.0	92.3	93.7	94.4	95.1	94.8	95.8	95.7	95.8	96.0	**93.4**

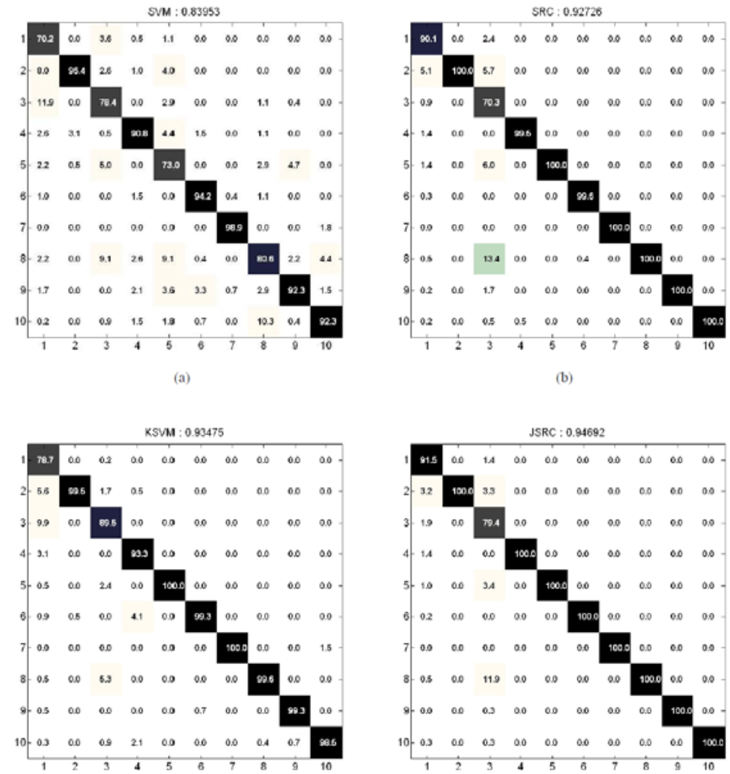
- Effect of the angle size between the views

AZIMUTH PERSISTENCE RESULTS (%) WITH DIFFERENT VIEW STEPS ($d = 256$, $M = 3$, $C = 3$, $N = 698$).

View-step size	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
SVM	76.4	77.1	79.6	78.0	79.3	81.4	79.9	79.7	80.4	81.0	81.1	81.2	80.7	80.5	80.4	80.2	80.7
SRC	86.3	87.5	87.6	87.4	87.9	87.5	88.3	89.3	87.3	88.1	86.5	86.4	87.4	85.6	87.8	85.2	85.6
KSVM	87.0	88.0	89.2	88.7	89.7	89.8	89.8	90.2	90.0	90.9	90.8	**91.3**	90.6	90.6	**90.9**	90.8	90.3
JSRC	**89.5**	**89.9**	**90.0**	**90.8**	**90.9**	**91.4**	**91.8**	**92.5**	**91.6**	**91.9**	**91.5**	90.9	**91.8**	**90.7**	90.8	**91.1**	**90.8**

Experimental Results on Multi-View Target Classification

- 10 class classification results using M=3 views with dictionary of size N=2747 tested on 15 degree depression



MULTI-VIEW ATR ACCURACY (%) ON MSTAR WITH DIFFERENT FEATURE DIMENSIONS ($M = 3$, $C = 10$, $N = 2747$)

Dims. (d)	128	256	512	1024	Avg.
SVM	65.1	70.6	76.9	84.0	74.15
SRC	73.0	84.7	90.1	92.7	85.13
KSVM	**82.5**	**87.8**	90.4	93.5	88.55
JSRC	79.7	87.7	**92.4**	**94.7**	**88.63**

Multi-Pose Face Recognition

- Scenarios where we have multiple poses of the same face as input to the classifier.
- UMIST database consists of 564 images of 20 individuals with a range of poses.
- Randomly select 10 poses for each individual to construct the dictionary.

Figure 8. Sample images from UMIST database for a single subject with varying poses.

Table 2. Multi-instance face recognition accuracy (%) on UMIST.

Algorithm	2 Views	3 Views	4 Views	Avg.
MSM [6]	94	95	96	95.0
GPH [11]	**96**	**97**	98	97.0
AFH [4]	94	96	97	95.7
MTJSRC [19]	95	96	98	96.3
JDSRC	**96**	**97**	**99**	**97.3**

Conclusions

- Formulated target and object recognition as joint sparsity underdetermined regression problem.
- Investigated the effect single vs multiple measurements
- Included the idea of joint structured sparsity prior into the regularization part of the optimization
- Investigated performance of multiple measurements on classification performance on several data bases.

THANK YOU