STRUCTURED SPARSE PRIORS FOR IMAGE CLASSIFICATION

Umamahesh Srinivas†, Yuanming Suo*, Minh Dao*, Vishal Monga†, Trac D. Tran*

†Department of Electrical Engineering, Pennsylvania State University, University Park, PA, USA
*Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA

Background: Sparsity in signal representation

- Compressive sensing (CS) for signal recovery

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_{0/1} \text{ subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

- Model-based compressive sensing
  1. Wavelet coefficients modeled as connected trees

$$\boldsymbol{x}_K^T = \arg\min_{\bar{\boldsymbol{x}}} \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2 \text{ s.t. } \bar{\boldsymbol{x}} \in \mathcal{T}_K$$

  2. Probabilistic prior models

$$\max_{\boldsymbol{x}} f(\boldsymbol{x}) \text{ subject to } \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 < \epsilon$$

  3. Laplacian MAP estimation

$$\arg\max_{\boldsymbol{x}} \exp\left\{-\lambda \sum_i |x_i|\right\} \Leftrightarrow \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1$$

Sparse representation-based classification (SRC)

- Use CS analytical framework → design class-specific dictionaries
- Assumption: New image lies in linear span of training from the same class

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

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

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

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

$$\boldsymbol{x} = \left[\boldsymbol{0}^T \ \ldots\ \boldsymbol{0}^T\ \boldsymbol{x}_i^T\ \boldsymbol{0}^T\ \ldots\ \boldsymbol{0}^T\right]^T$$

- 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:

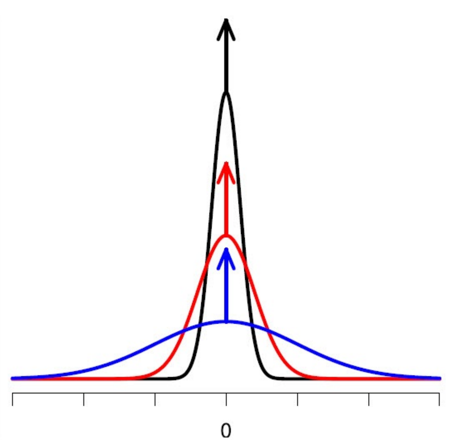
$$\text{identity}(\boldsymbol{y}) = \arg\min_i \|\boldsymbol{y} - \boldsymbol{A}\delta_i(\hat{\boldsymbol{x}})\|_2$$

Contribution: SSPIC

- **Challenge:**
  - Abundant training necessary for sparse linear model to be valid
- **Solution:** Model-based extension to SRC
  - Design class-specific priors to capture discriminative sparse structure
- What priors to use?
- Spike-and-slab prior

$$x \sim (1 - \gamma)\delta_0 + \gamma p(x)$$

- Gold standard for sparse Bayesian inference (Titsias, NIPS 2011)



Set-theoretic interpretation

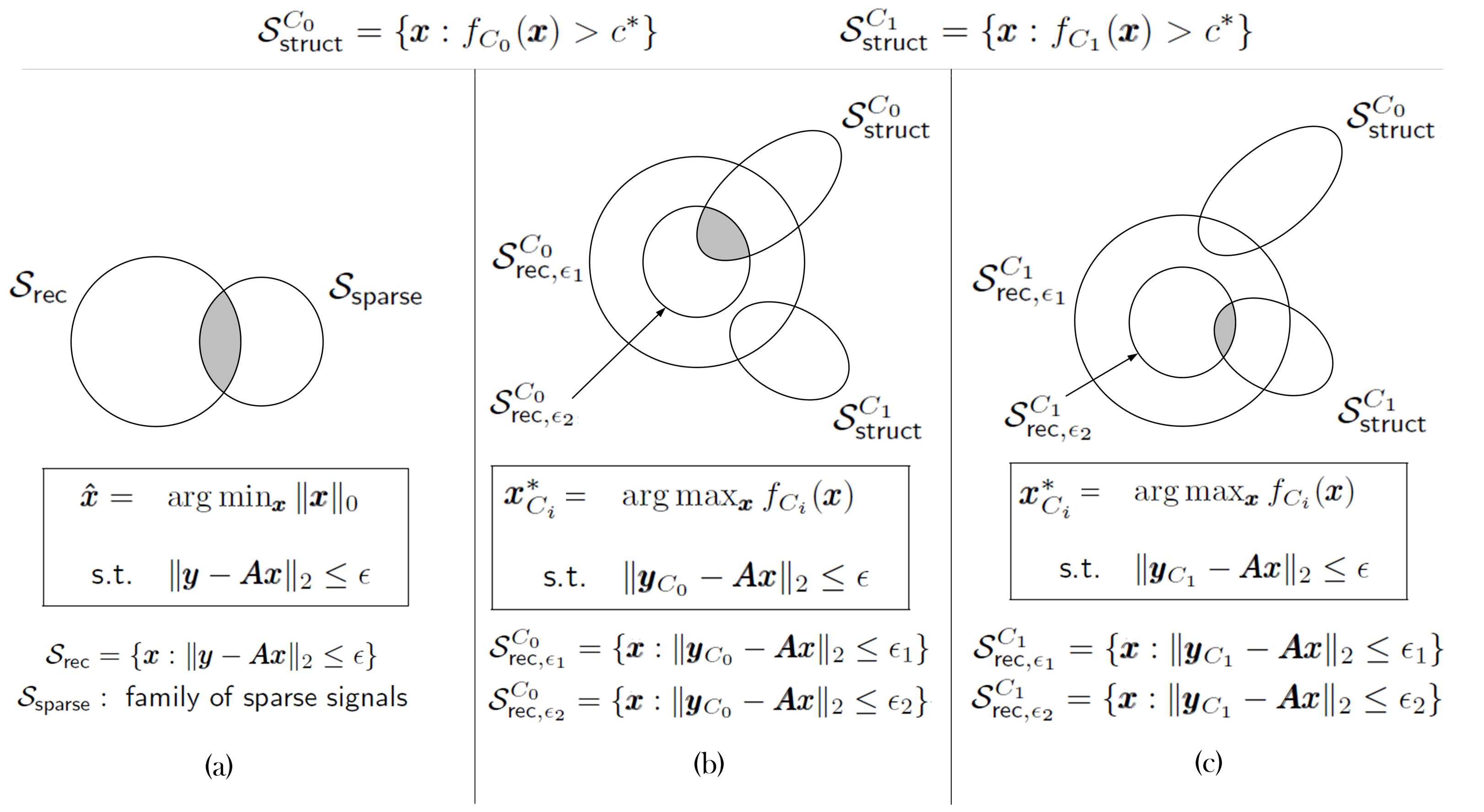
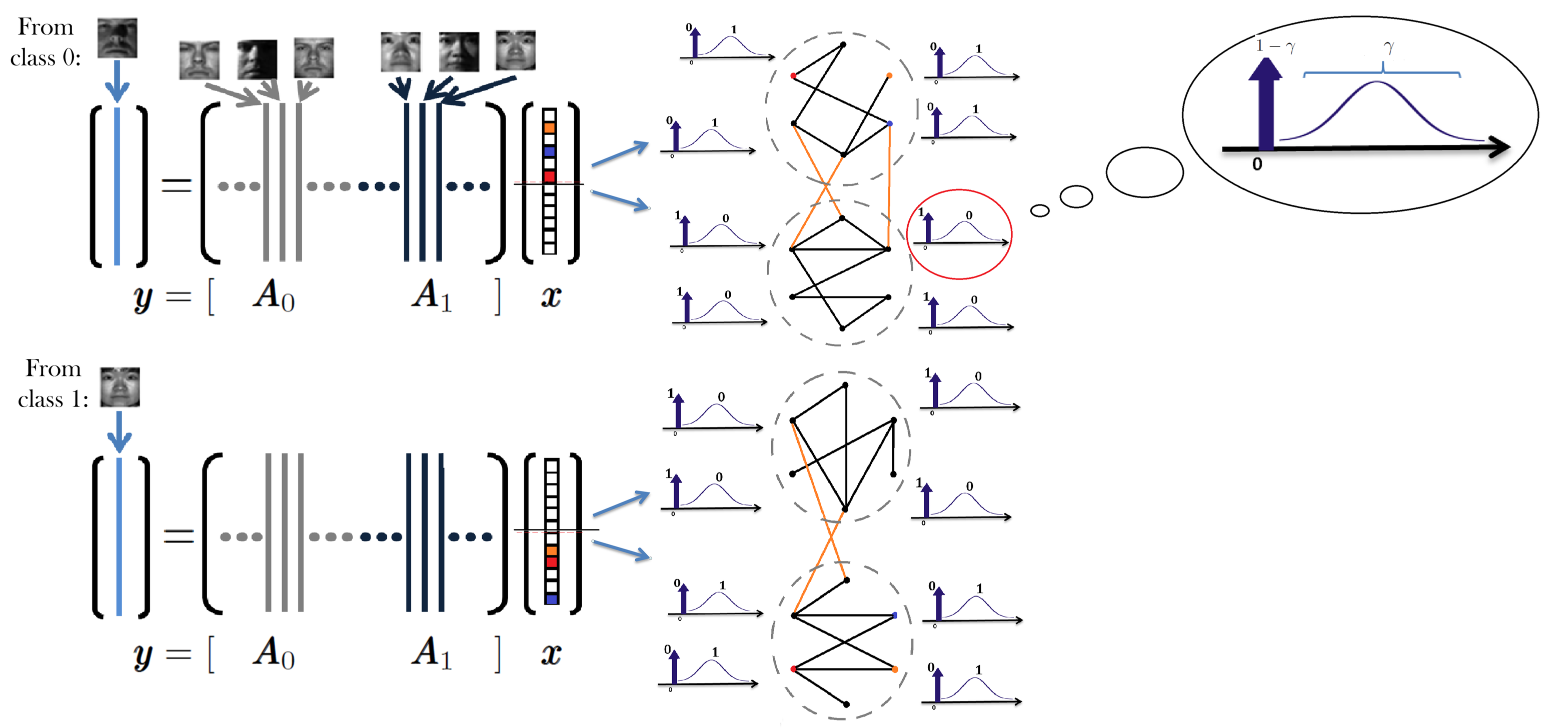


Illustration: Face recognition using spike-and-slab priors



Analytical formulation

- Bayesian set-up:

$$\boldsymbol{y}|\boldsymbol{A}, \boldsymbol{x}, \gamma, \sigma^2 \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{x}, \sigma^2\boldsymbol{I})$$
$$x_i|\sigma^2, \gamma_i, \lambda \sim \gamma_i\mathcal{N}(0, \sigma^2\lambda^{-1}) + (1 - \gamma_i)\mathbb{I}(x_i = 0)$$
$$\sigma^2|\tau_1, \tau_2 \sim \Gamma^{-1}(\tau_1, \tau_2)$$
$$\gamma_i|\kappa_i \sim \text{Bernoulli}(\kappa_i),\ i = 1, \ldots, n.$$

- MAP estimation

$$f(\boldsymbol{x}, \gamma, \sigma^2|\boldsymbol{A}, \boldsymbol{y}, \lambda, \tau_1, \tau_2, \boldsymbol{\kappa}) \propto f(\boldsymbol{y}|\boldsymbol{A}, \boldsymbol{x}, \gamma, \sigma^2) f(\boldsymbol{x}|\gamma, \sigma^2, \lambda) f(\sigma^2|\tau_1, \tau_2) f(\gamma|\boldsymbol{\kappa})$$

$$(\boldsymbol{x}^*, \gamma^*, \sigma^{2*}) = \arg\min \frac{1}{\sigma^2}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) + m\log\sigma^2 + \frac{\boldsymbol{x}^T\boldsymbol{x}}{\sigma^2\lambda^{-1}} + \left(\sum_{i=1}^n \gamma_i\right)\log\left(\frac{2\pi\sigma^2}{\lambda}\right) + \frac{2\tau_2}{\sigma^2} + 2(\tau_1 + 1)\log\sigma^2 + \sum_{i=1}^n \gamma_i \log\left(\frac{(1 - \kappa_i)^2}{\kappa_i^2}\right)$$

- Tractability of optimization problem → choose single κ per class

Summary and observations

- Modified optimization problem

$$L(\boldsymbol{x}, \gamma, \sigma^2) = \frac{1}{\sigma^2}\left\{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_2^2 + \rho_{\sigma^2,\lambda,\kappa}\|\boldsymbol{x}\|_0 + (m + 2\tau_1 + 2)\sigma^2\log\sigma^2 + 2\tau_2\right\}$$

- Multiple per-class problems

$$\boldsymbol{x}^*_{C_i} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda_{C_i}\|\boldsymbol{x}\|_2^2 + \rho_{C_i,\sigma^2,\lambda,\kappa}\|\boldsymbol{x}\|_0$$

- Comparison with SRC
  1. Sparse structure enforced by a weighted combination of ℓ_0 and ℓ_2-terms
  2. Class-specific weights on ℓ_0 and ℓ_2-terms → more discriminative compared to class-specific dictionaries in SRC
- Practical benefit: Use of well designed class-specific priors alleviates burden on number of training images required

Problem I: Face recognition

Table: Overall recognition rates: Extended Yale B database

Method	Recognition rate (%)
SSPIC	**97.3**
SRC	97.1
Eigen-NS	89.5
Eigen-SVM	91.9
Fisher-NS	84.7
Fisher-SVM	92.6

Table: Disguise: AR database

Method	Recognition rate (%) Sunglasses	Recognition rate (%) Scarves
SSPIC	**95.1**	**91.6**
SRC	93.5	90.1
Eigen-NS	47.2	29.6
Eigen-SVM	53.5	34.5
Fisher-NS	57.9	41.7
Fisher-SVM	61.7	43.6

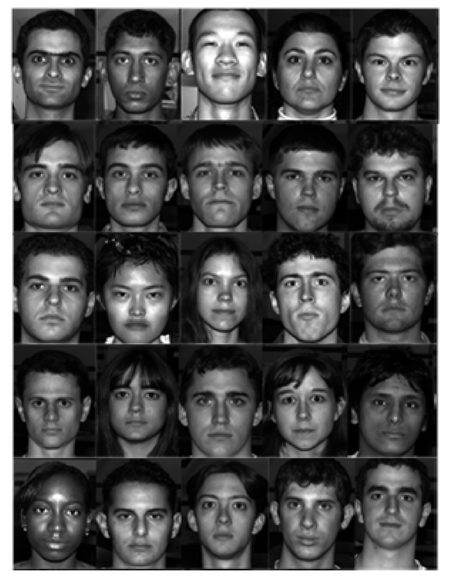


Figure: Sample images: Extended Yale B.

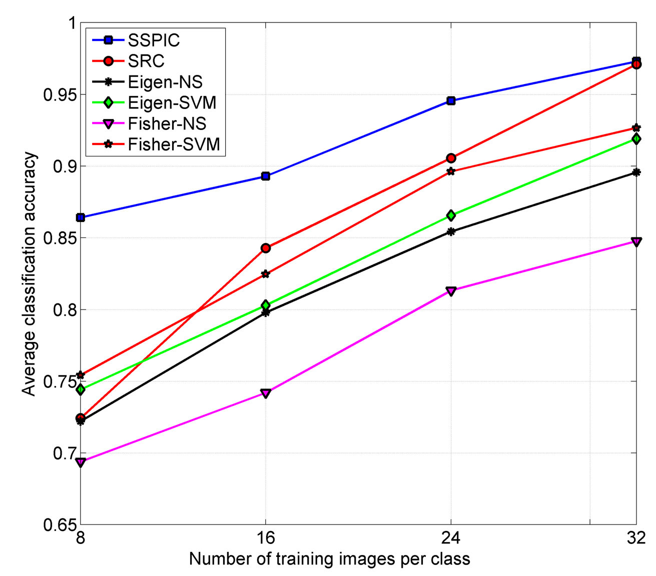


Figure: Classification variation with training set size: Extended Yale B.

Problem II: Object categorization (Caltech-101 dataset)

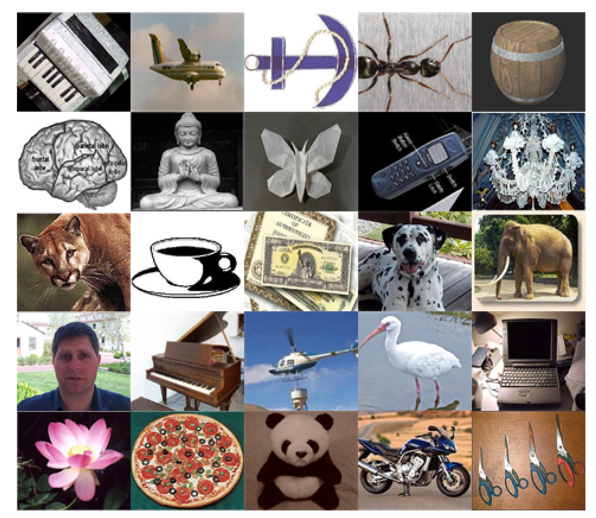


Figure: Sample images.

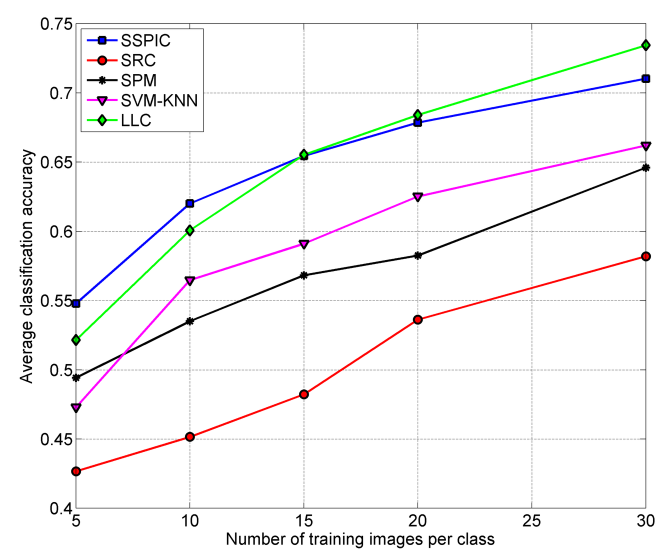


Figure: Classification variation with training set size.

Acknowledgment: ONR Grant N00014-12-1-0765

Email: vmonga@engr.psu.edu, Web: http://signal.ee.psu.edu/