Structured Analysis Dictionary Learning for Image Classification

Wen Tang, Ashkan Panahi, Hamid Krim, Liyi Dai*

North Carolina State University, USA
*Army Research Office, USA

International Conference on Acoustics, Speech and Signal Processing
April 17, 2018

Overview

- **Introduction**
- **Structured Analysis Dictionary Learning**
  - **Conventional ADL**
  - **Mitigating Inter-Class Feature Interference**
    - ❖ **Structural Mapping of Sparse Representation**
    - ❖ **Minimal Classification Error**
- **Experiments and Results**
- **Conclusion**

Task

Task-driven DL

Synthesis Dictionary Learning

$$\min_{\Omega,U} \frac{1}{2}\|X-\Omega U\|_2^2+\lambda\|U\|_1$$

1. Learn class-specific dictionaries

$$\min_{\Omega_i,A_i} \frac{1}{2}\|X_i-\Omega_i U_i\|_2^2+\lambda\|U_i\|_1, \qquad \forall\, i=1,\ldots,C$$

2. Jointly learn a universal dictionary and a multiclass classifier

$$\min_{\Omega,A,W} \frac{1}{2}\|X-\Omega U\|_2^2+\lambda_1\|U\|_1+\lambda_2\|L-WU\|_2^2$$

Task-driven Analysis DL

Analysis Dictionary Learning

$$\min_{\Omega,U} \frac{1}{2}\|U-\Omega X\|_2^2+\lambda\|U\|_1$$

Analysis K-SVD, Sparse Null Space (SNS) pursuit

1. [Shekhar et al., 2014]: ADL + SVM

2. [Guo et al., 2016]: topological structures & discriminative labels & ADL +KNN.

Our Work

- Based on ADL framework:
  - A structural mapping:
    - Sparse representations are more consistent
  - Classification error feedback:
    - Discriminative multiclass classifier jointly learned
- Efficiently solved by Linearized ADM
- Comparable or better accuracies with extremely fast testing time

Overview

- Introduction
- **Structured Analysis Dictionary Learning**
  - **Conventional ADL**
  - **Mitigating Inter-Class Feature Interference**
    - ❖ **Structural Mapping of Sparse Representation**
    - ❖ **Minimal Classification Error**
- Experiments and Results
- Conclusion

Conventional ADL

Analysis Dictionary

Sparse Representation

Data

$$\min_{\Omega,U} \frac{1}{2} \|U - \Omega X\|_2^2 + \lambda \|U\|_1$$

$$s.t. \quad \Omega \in \mathcal{W}$$

Non-trivial Solution

Classification performance is **poor!**

Structural Mapping of Sparse Representation

Transform Matrix

Structured Representation

$$H = QU + \varepsilon_1$$

Tolerance

Example:

$$H = \begin{array}{c} \begin{matrix} h_1^1 & h_2^1 & h_3^1 & h_4^2 & h_5^2 \end{matrix} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

Diagonal Block Matrix

Minimal Classification Error

Regression Classifier

$$L = WQU + \varepsilon_2$$

Classes Labels

Tolerance

Example:

$$L = \begin{array}{c} \begin{matrix} c_1 & c_1 & c_1 & c_2 & c_2 \end{matrix} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

Structured Analysis Dictionary Learning

Analysis Dictionary

Data

Sparse Representation

$$\underset{\substack{\Omega, U, Q, W, \\ \varepsilon_1, \varepsilon_2}}{\operatorname{argmin}} \frac{1}{2}\|U - \Omega X\|_F^2 + \lambda_1 \|U\|_1 + \frac{\rho_1}{2}\|\varepsilon_1\|_2^2 + \frac{\rho_2}{2}\|\varepsilon_2\|_2^2$$

Transform Matrix

Tolerance

$$s.t.\ H = QU + \varepsilon_1,$$

Structured Representation

$$L = W(QU) + \varepsilon_2,$$

Class Label

Regression Classifier

$$\left\|\omega_i^T\right\|_2^2 = 1;\ \forall i = 1, \ldots, r.$$

Atoms in Dictionary

Augmented Lagrangian

$$L\left(\Omega, U, Q, W, Y^{(1)}, Y^{(2)}, \mu\right) = \frac{1}{2}\|U - \Omega X\|_F^2 + \lambda_1 \|U\|_1$$
$$+ \lambda_2 < Y^{(1)}, H - QU > + \lambda_3 < Y^{(2)}, L - WQU >$$
$$+ \frac{\mu}{2}\|H - QU\|_F^2 + \frac{\mu}{2}\|L - WQU\|_F^2$$

Tuning Parameters

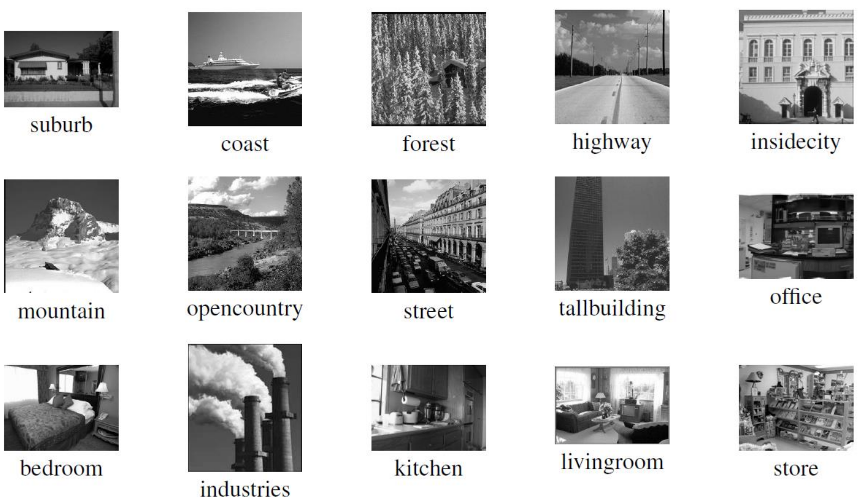
Overview

- Introduction
- Structured Analysis Dictionary Learning
  - Conventional ADL
  - Mitigating Inter-Class Feature Interference
    - Structural Mapping of Sparse Representation
    - Minimal Classification Error
- **Experiments and Results**
- Conclusion

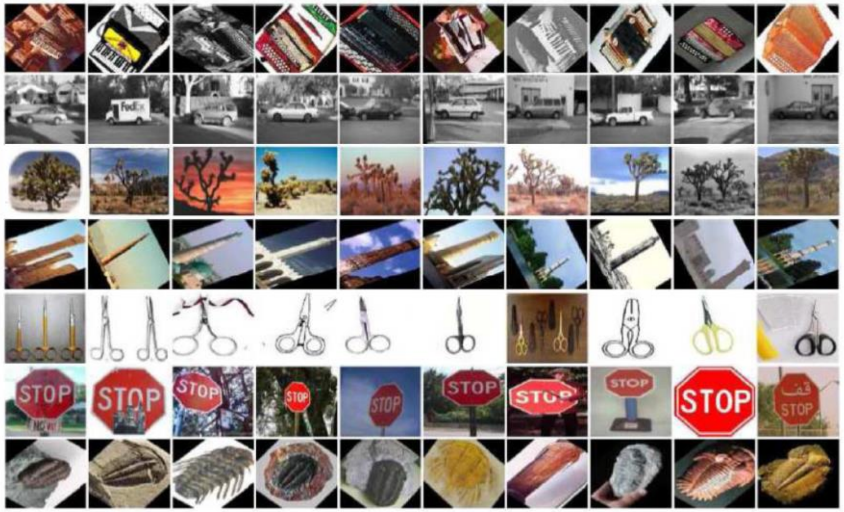
Evaluated Database

Extended YaleB

AR



Scene15



Caltech101

Parameter Settings

- Parameters chosen by 10-fold cross validation.

State-of-the-art Methods

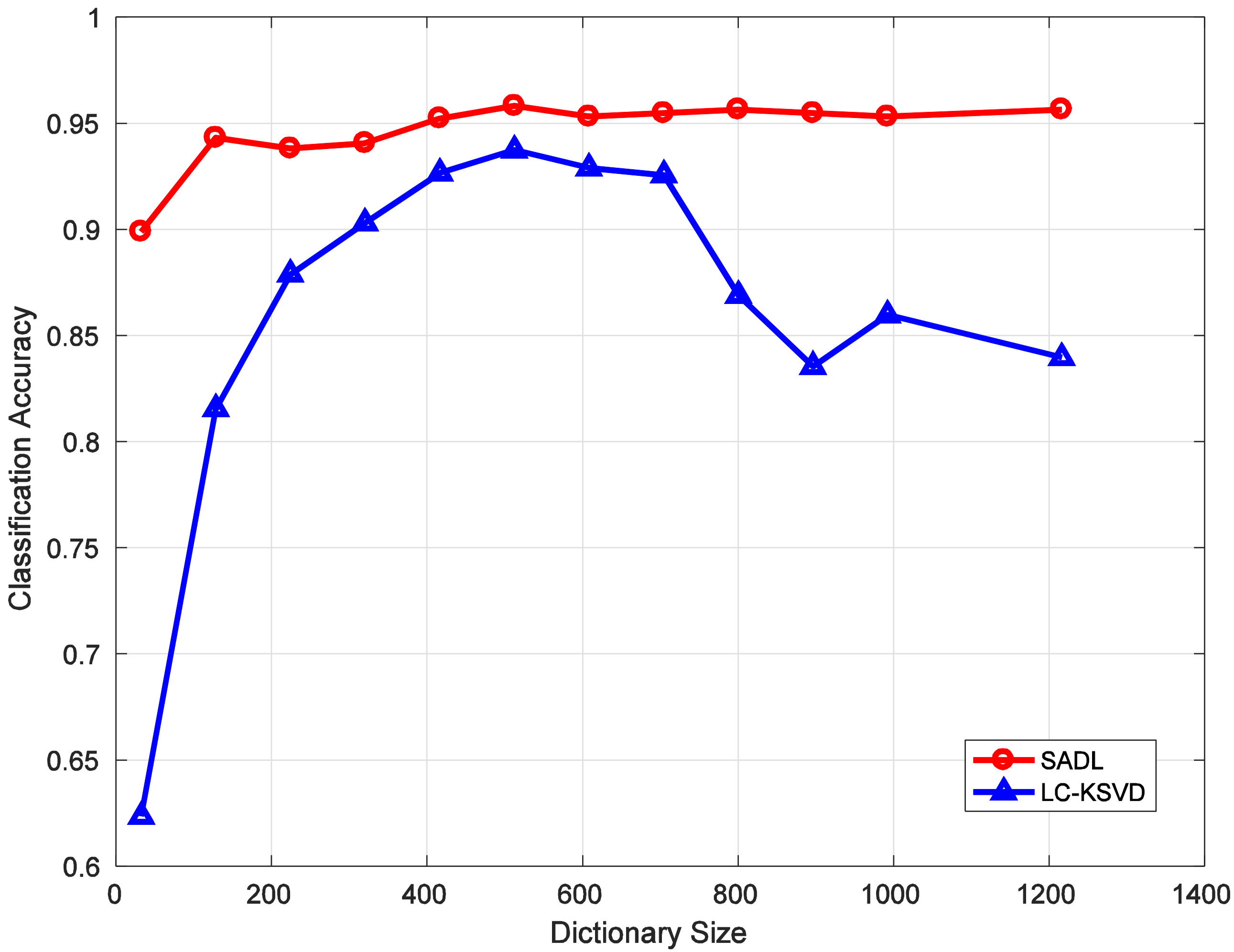
1. ADL+SVM: sparse representations learned by ADL and classified by SVM.
2. SRC: sparse representations learned by the dictionary composed of training images.
3. LC-KSVD: forces each category labels to be consistent.

Extended YaleB

Methods	Classification Accuracy(%)	Training Time(s)	Testing Time(s)
ADL+SVM	82.91%	91.78	1.13×10^{-3}
SRC	80.5%	$No\ Need$	3.74×10^{-1}
LC-KSVD	94.56% $(\mathbf{95\%})$	234.67	1.63×10^{-2}
SADL	94.91%	$\mathbf{51.29}$	$\mathbf{2.72 \times 10^{-6}}$

*95% was reported in the original paper of LC-KSVD.

Extended YaleB Dataset

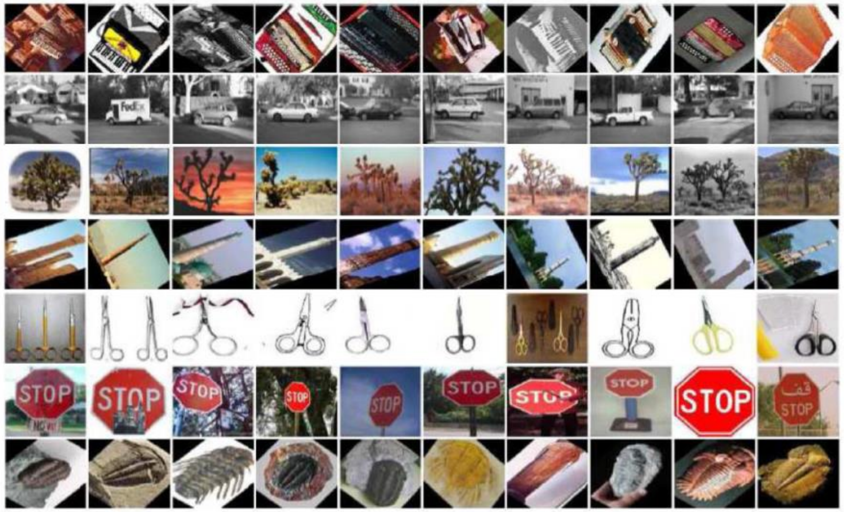


AR Face

Methods	Classification Accuracy(%)	Training Time(s)	Testing Time(s)
ADL+SVM	90.40%	218.54	9.10×10^{-3}
SRC	66.50%	*No Need*	5.25×10^{-2}
LC-KSVD	87.78% (93.7%)	244.52	1.42×10^{-2}
SADL	**95.08%**	**89.13**	$\mathbf{3.67 \times 10^{-6}}$

*93.7% was reported in the original paper of LC-KSVD.

Caltech101



Methods	Classification Accuracy(%)	Training Time(s)	Testing Time(s)
ADL+SVM	54.93%	$\mathbf{447.80}$	7.75×10^{-3}
SRC	67.70%	*No Need*	4.34×10^{-1}
LC-KSVD	71.79%	487.61	1.35×10^{-2}
SADL	$\mathbf{72.36\%}$	773.66	$\mathbf{8.10 \times 10^{-6}}$

Scene 15

Methods	Classification Accuracy(%)	Training Time(s)	Testing Time(s)
ADL+SVM	49.35%	$\mathbf{110.47}$	1.14×10^{-4}
SRC	91.80%	*No Need*	4.06×10^{-1}
LC-KSVD	$\mathbf{98.83\%}$ (92.9%)	270.93	1.26×10^{-2}
SADL	98.16%	121.02	$\mathbf{9.23 \times 10^{-6}}$

*92.9% was reported in the original paper of LC-KSVD.

Conclusion

- A structural mapping and a classification fidelity are included.
- Optimization problem efficiently solved by linearized ADM.
- Performances are comparable or better and more stable.
- Thousands of times faster for testing.

Thank you!