Joint Weighted Dictionary Learning and Classifier Training for Robust Biometric Recognition

Rahman Khorsandi, Ali Taalimi, Mohamed Abdel-Mottaleb, Hairong Qi

University of Miami, University of Tennessee, Knoxville

{r.khorsandi,mottaleb@umiami.edu}

{ataalimi, hqi}@utk.edu

Outline

- **Background**
- **Motivation**
- **Contribution**
- **Experiment**

Background

Sparse Representation-based Classification

- **Training Phase:**

A. The training dataset:

$$\mathbf{Y}_c \in \mathcal{R}^{m \times n_c} = \{\mathbf{y}_{i,c} \,|\, i \in \{1, \cdots, n_c\}\}$$
$$\mathbf{Y} \in \mathcal{R}^{m \times N} = \{\mathbf{Y}_c \,|\, c \in \{1, \cdots, C\}\}$$

J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(2):210–227, Feb 2009.

Background

Sparse Representation-based Classification

- **Training Phase:**

B. Make the dictionary:

1. Fixed

$$\mathbf{D}_c = [\mathbf{y}_{1,c}, \cdots, \mathbf{y}_{N_c,c}], \quad \mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_C]$$

J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(2):210–227, Feb 2009.

Background

Sparse Representation-based Classification

- **Training Phase:**

B. Make the dictionary:

1. Fixed:
2. Learning: $$\min_{\mathbf{D}} \triangleq \frac{1}{N} \sum_i \mathcal{L}_u(\mathbf{x}_i, \mathbf{D})$$

$$\mathcal{L}_u(\mathbf{x}_i, \mathbf{D}) \triangleq \min_{\mathbf{x}_i \in \mathbb{R}^p} \left\| \mathbf{y}_{i,c} - \mathbf{D}_c \mathbf{x}_{i,c} \right\|_2^2 + \lambda_1 \left\| \mathbf{x}_{i,c} \right\|_1 + \lambda_2 \left\| \mathbf{x}_{i,c} \right\|_2$$

Background

Testing Phase

- **Dictionary D from training phase:**

A. Reconstruct test signal.

$$\min_{\mathbf{x}_t^* \in \mathbb{R}^p} \left\|\mathbf{y}_t - \mathbf{D}\mathbf{x}_t^*\right\|_2^2 + \lambda_1 \left\|\mathbf{x}_t^*\right\|_1 + \lambda_2 \left\|\mathbf{x}_t^*\right\|_2$$

$$\delta_C(\mathbf{x}) = [\mathbf{0}^T, \cdots, \mathbf{0}^T, \mathbf{x}_c^T, \cdots, \mathbf{0}^T]^T$$

$$c^* = \min_c \left\|\mathbf{y}_t - \mathbf{D}_c \delta_c(\mathbf{x}_t)\right\|_{\ell_2}$$

B. Feature extraction.

$$\min_{\mathbf{W}} \sum_{i=1}^{N} \left\|\mathbf{h}_i - \mathbf{W}\mathbf{x}_i\right\|_2^2 + \frac{\nu}{2} \left\|\mathbf{W}\right\|_F^2$$

Motivation

Sparse Representation-based Classification

- **Issues:**

  A. Apply unsupervised dictionary for discriminative task.

  B. Atoms are not required to be uncorrelated.

  C. Train an independent classifier.

J.Mairal,F.Bach and J.Ponce. Task-driven dictionary learning. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(4):791– 804, April 2012.

Contribution

Learn classifier
wrt the current dictionary

Task Driven Optimization

Minimize correlation between atoms
Dictionary that is discriminative and reconstructive

Weighted Dictionary Learning

Bi-level Optimization

J.Mairal,F.Bach and J.Ponce. Task-driven dictionary learning. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(4):791– 804, April 2012.

M. Yang, D. Dai, L. Shen, and L. Van Gool. Latent dictionary learning for sparse representation based classification. In Computer Vision and Pattern Recognition (CVPR), 2014.

Contribution

Task driven dictionary learning

- **Joint estimation of dictionary and classifier:**

$$\min_{\mathbf{D},\mathbf{W}} \sum_{i=1}^{N} \left\| \mathbf{h}_i - \mathbf{W}\mathbf{x}_i^*(y_i, D) \right\|_2^2 + \frac{\nu}{2} \| \mathbf{W} \|_F^2$$

$$\min_{\mathbf{x}_i,\mathbf{D}} \left\| \mathbf{y}_i - \mathbf{D}\mathbf{x}_i \right\|_2^2 + \lambda \left\| \mathbf{x}_i \right\|_1 \quad \forall i$$

Mairal, J., F. Bach, and J. Ponce (2012) "Task-driven dictionary learning," IEEE Trans. Pattern Anal. Mach. Intell., 34(4), pp. 791–804.

Contribution

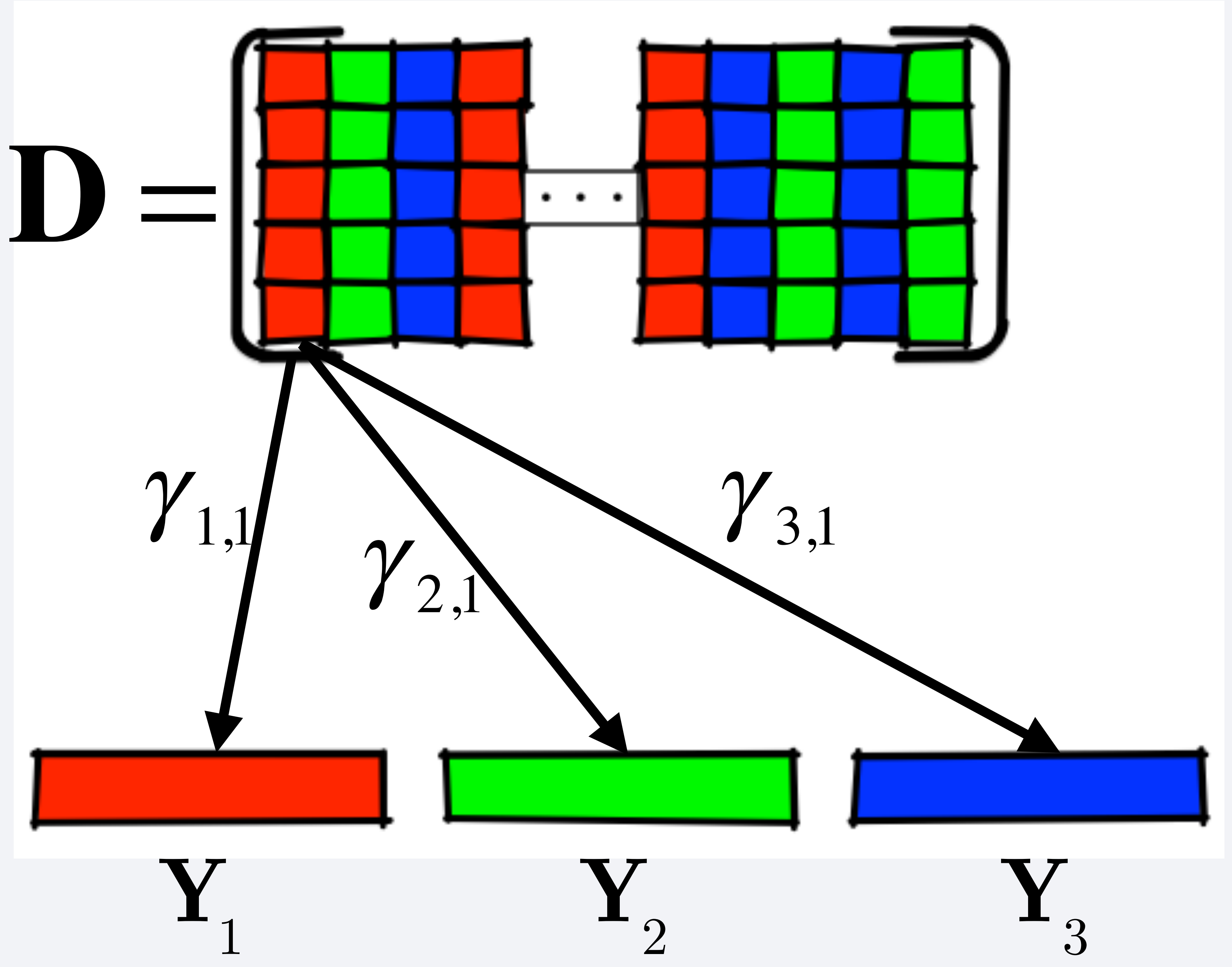
Weighted Dictionary Learning

- Correlation between atoms and c-th class:

$$\gamma_c = [\gamma_{c,1}, \gamma_{c,2}, \cdots, \gamma_{c,p}]^T,$$

$$\gamma_{c,i} \geq 0, \forall i \in \{1, \cdots, p\}$$

$$\sum_{i=1}^{p} \gamma_{c,i} = \sum_{i=1}^{p} \gamma_{c',i}$$



$$\mathrm{corr}(d_i, d_j) \uparrow \Leftrightarrow (\gamma_{c,j}\gamma_{l,i}) \downarrow$$

M. Yang, D. Dai, L. Shen, and L. Van Gool. Latent dictionary learning for sparse representation based classification. In Computer Vision and Pattern Recognition (CVPR), 2014.

Contribution

Weighted Dictionary Learning

$$\min_{\mathbf{D},\gamma_c,\mathbf{X}} \sum_{c=1}^{C} \left\| \mathbf{Y}_c - \mathbf{D}\,diag(\gamma_c)\mathbf{X}_c \right\|_F^2 + \boldsymbol{\lambda}_1 \left\| \mathbf{X}_c \right\|_1$$

$$+\lambda_2 \sum_{c=1}^{C} \sum_{l \neq c} \sum_{i=1}^{p} \sum_{j \neq i} \gamma_{c,j} (\mathbf{d}_j^T \mathbf{d}_i)^2 \gamma_{l,i}$$

$$\gamma_{c,i} \geq 0 \text{ and } \sum_{i=1}^{p} \gamma_{c,i} = \sum_{i=1}^{p} \gamma_{l,i}, \ \forall c, l$$

- **Reconstruct data as:**

$$\mathbf{Y}_c \approx \mathbf{D}\,diag(\gamma_c)\mathbf{X}_c$$

M. Yang, D. Dai, L. Shen, and L. Van Gool. Latent dictionary learning for sparse representation based classification. In Computer Vision and Pattern Recognition (CVPR), 2014.

Contribution

Joint Weighted Dictionary Learning and Classifier Training

$$\min_{\mathbf{D},\mathbf{W}} \sum_{i=1}^{N} \left\| \mathbf{h}_i - \mathbf{W}\mathbf{x}_i^*(\boldsymbol{y}_i, \mathbf{D}) \right\|_2^2 + \frac{\nu}{2} \| \mathbf{W} \|_F^2$$

← Task Driven

$$\min_{\mathbf{D},\boldsymbol{\Gamma},\mathbf{X}} \sum_{c=1}^{C} \left\| \mathbf{Y}_c - \mathbf{D}\, diag(\gamma_c)\mathbf{X}_c \right\|_F^2 + \boldsymbol{\lambda}_1 \left\| \mathbf{X}_c \right\|_1$$

$$+\lambda_2 \sum_{c=1}^{C} \sum_{l \neq c} \sum_{i=1}^{p} \sum_{j \neq i} \gamma_{c,j} (\mathbf{d}_j^T \mathbf{d}_i)^2 \gamma_{l,i}$$

$$\gamma_{c,i} \geq 0 \text{ and } \sum_{i=1}^{p} \gamma_{c,i} = \sum_{i=1}^{p} \gamma_{l,i}, \forall c, l$$

← Weighted Dictionary Learning

Optimization

First Step

Constant γ_c

solve for $\mathbf{D}$ and $\mathbf{X}$

→

Optimization

First Step

Constant γ_c

solve for $\mathbf{D}_c$ and $\mathbf{X}_c$

$$\min_{\mathbf{X}_c} \frac{1}{2}\left\|\mathbf{Y}_c - \mathbf{D}\,diag(\gamma_c)\mathbf{X}_c\right\|_F^2 + \lambda_2\left\|\mathbf{X}_c\right\|_1$$

$$\min_{\mathbf{D}} \sum_{c=1}^{C}\left\|\mathbf{Y}_c - \mathbf{D}\,diag(\gamma_c)\mathbf{X}_c\right\|_F^2 +$$

$$\lambda_2\sum_{c=1}^{C}\sum_{l\neq c}\sum_{i=1}^{p}\sum_{j\neq i}\gamma_{c,j}(\mathbf{d}_j^T\mathbf{d}_i)^2\gamma_{l,i}$$

L. Rosasco, A. Verri, M. Santoro, S. Mosci, and S. Villa. Iterative projection methods for structured sparsity regularization. 2009.

M. Yang, D. Dai, L. Shen, and L. Van Gool. Latent dictionary learning for sparse representation based classification. In Computer Vision and Pattern Recognition (CVPR), 2014.

Optimization

Second Step

Constant $\mathbf{D}$ and $\mathbf{X}$

solve for γ_c

$$\min_{\gamma_c} \left\| \mathbf{Y}_c - \mathbf{D}\, diag(\gamma_c)\mathbf{X}_c \right\|_F^2 + 2\lambda_2 \sum_{i=1}^{p} \gamma_{c,i} \sum_{j \neq i} (\mathbf{d}_j^{\intercal}\mathbf{d}_i)^2 \sum_{l \neq c} \gamma_{l,j}$$

Third Step

Constant $\mathbf{X}$

solve for $\mathbf{W}$

$$\min_{\mathbf{W}} \sum_{i=1}^{N} \left\| \mathbf{h}_i - \mathbf{W}\mathbf{x}_i \right\|_2^2 + \frac{\nu}{2} \left\| \mathbf{W} \right\|_F^2$$

Z. Jiang, Z. Lin, and L. S. Davis. Label consistent k-svd: learning a discriminative dictionary for recognition. Pattern Analysis and Machine Intelligence, 2013.

L. Rosasco, A. Verri, M. Santoro, S. Mosci, and S. Villa. Iterative projection methods for structured sparsity regularization. 2009.

Mairal, J., F. Bach, and J. Ponce (2012) "Task-driven dictionary learning," IEEE Trans. Pattern Anal. Mach. Intell., 34(4), pp. 791–804.

Testing phase

- **From Training:** $\mathbf{D}, \gamma_c$

$$\min_{\mathbf{x}_{t,c}} \left\|\mathbf{Y}_t - \mathbf{D}\,diag(\gamma_c)\mathbf{x}_{t,c}\right\|_F^2 + \boldsymbol{\lambda}_1 \left\|\mathbf{x}_{t,c}\right\|_1$$

$$+\lambda_2 \sum_{l \neq c} \sum_{i=1}^{p} \sum_{j \neq i} \gamma_{c,j} (\mathbf{d}_j^T \mathbf{d}_i)^2 \gamma_{l,i}$$

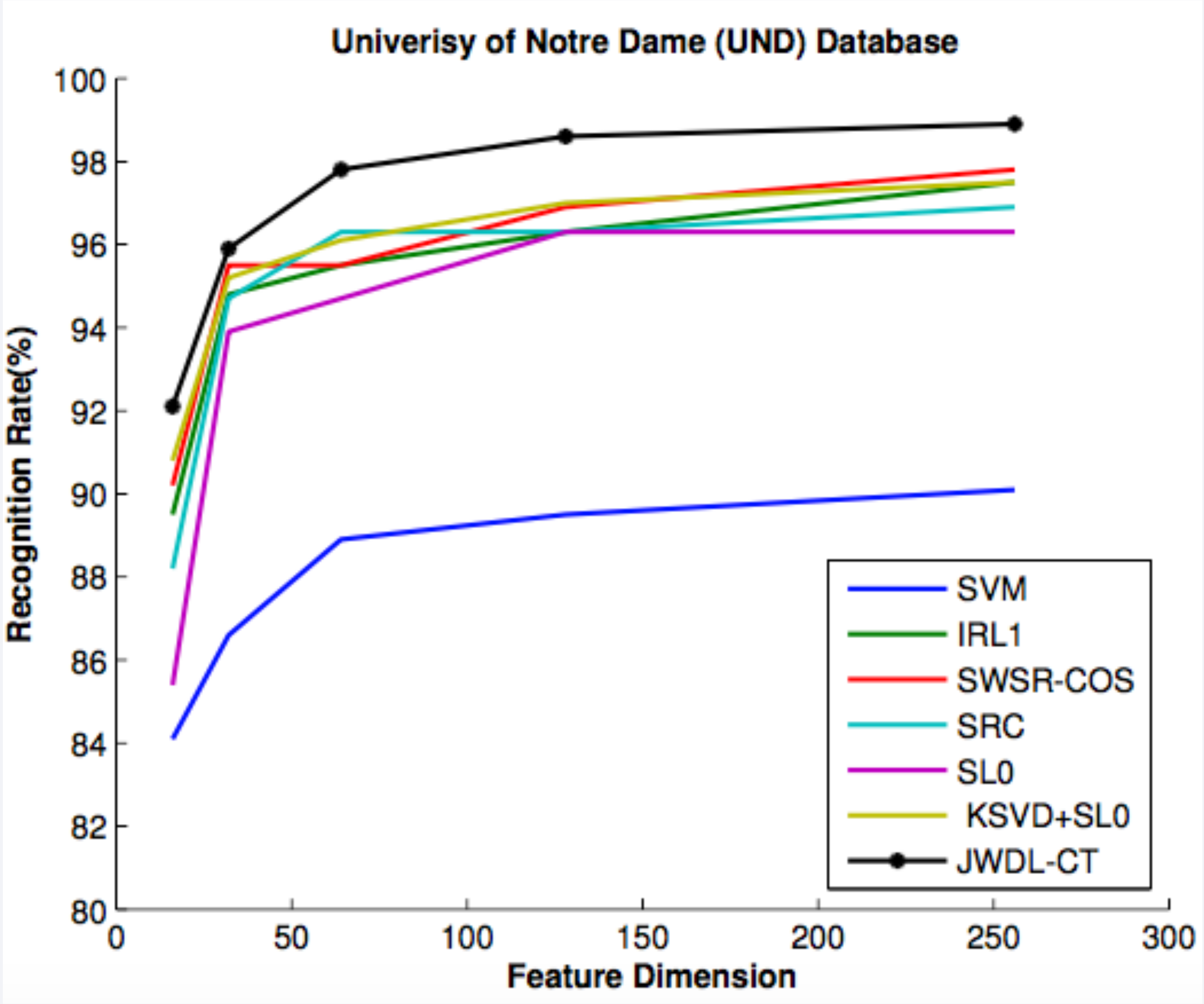
Weighted Dictionary Learning

$$\mathcal{E}_{t,c} = \| \mathbf{y}_t - \mathbf{D}\,diag(\gamma_c)\mathbf{x}_{t,c} \|_2^2 + \eta \| \mathbf{h}_c - \mathbf{W}\mathbf{x}_{t,c} \|_2^2$$

Quantitative Comparison

UND data set

A. 235 male and 169 female subjects.



Quantitative Comparison

WVU data set

- 5 atoms per subject.
- HoG feature.

Number of Atoms (per subject) in the Dictionary	5	7	9
NN	74.1%	77.3%	79.8%
SVM	72.8%	74.1%	76.5%
Adaboost	68.9%	72.3%	75.5%
SRC	65.1%	66.0%	68.2%
LDL	77.5%	78.1%	82.1%
FDDL	79.5%	82.3%	84.5%
JWDL-CT	83.6%	85.7%	88.2%

EAR RECOGNITION RATES ON WVU DATABASE.

Corruption Ratio	5%	10%	20%
NN	72.5%	71.1%	68.1%
SVM	71.6%	70.2%	68.0%
Adaboost	67.1%	65.9%	62.3%
SRC	64.1%	63.5%	62.2%
K-SVD + SL0	75.9%	73.6%	70.0%
FDDL	78.0%	76.6%	73.8%
JWDL-CT	82.9%	81.2%	79.7%

EAR RECOGNITION RATES UNDER DIFFERENT RATIOS OF RANDOM CORRUPTION.