FAST LOW-RANK SHARED DICTIONARY LEARNING FOR IMAGE CLASSIFICATION

國立中興大學
National ChungHsing University

Institute of Statistics &
Department of Applied Mathematics

$$\oint_{\text{NCHU}} f(\text{amath})\, dS_t$$

Huynh Huu Dinh

Email: huynhhuudinh@iuh.edu.vn

Fast Low-Rank Shared Dictionary Learning for Image Classification

2 Author(s) Tiep Huu Vu ; Vishal Monga View All Authors

Abstract:
Despite the fact that different objects possess distinct class-specific features, they also usually share common patterns. This observation has been exploited partially in a recently proposed dictionary learning framework by separating the particularity and the commonality (COPAR). Inspired by this, we propose a novel method to explicitly and simultaneously learn a set of common patterns as well as class-specific features for classification with more intuitive constraints. Our dictionary learning framework is hence characterized by both a shared dictionary and particular (class-specific) dictionaries. For the shared dictionary, we enforce a low-rank constraint, i.e., claim that its spanning subspace should have low dimension and the coefficients corresponding to this dictionary should be similar. For the particular dictionaries, we impose on them the well-known constraints stated in the Fisher discrimination dictionary learning (FDDL). Furthermore, we develop new fast and accurate algorithms to solve the subproblems in the learning step, accelerating its convergence. The said algorithms could also be applied to FDDL and its extensions. The efficiencies of these algorithms are theoretically and experimentally verified by comparing their complexities and running time with those of other well-known dictionary learning methods. Experimental results on widely used image data sets establish the advantages of our method over the state-of-the-art dictionary learning methods.

Published in: IEEE Transactions on Image Processing (Volume: 26 , Issue: 11 , Nov. 2017)

Page(s): 5160 - 5175

Date of Publication: 20 July 2017

ISSN Information:

INSPEC Accession Number: 17135638

DOI: 10.1109/TIP.2017.2729885

Publisher: IEEE

Sponsored by: IEEE Signal Processing Society

Purpose of paper

- Different objects possess distinct class-specific features.

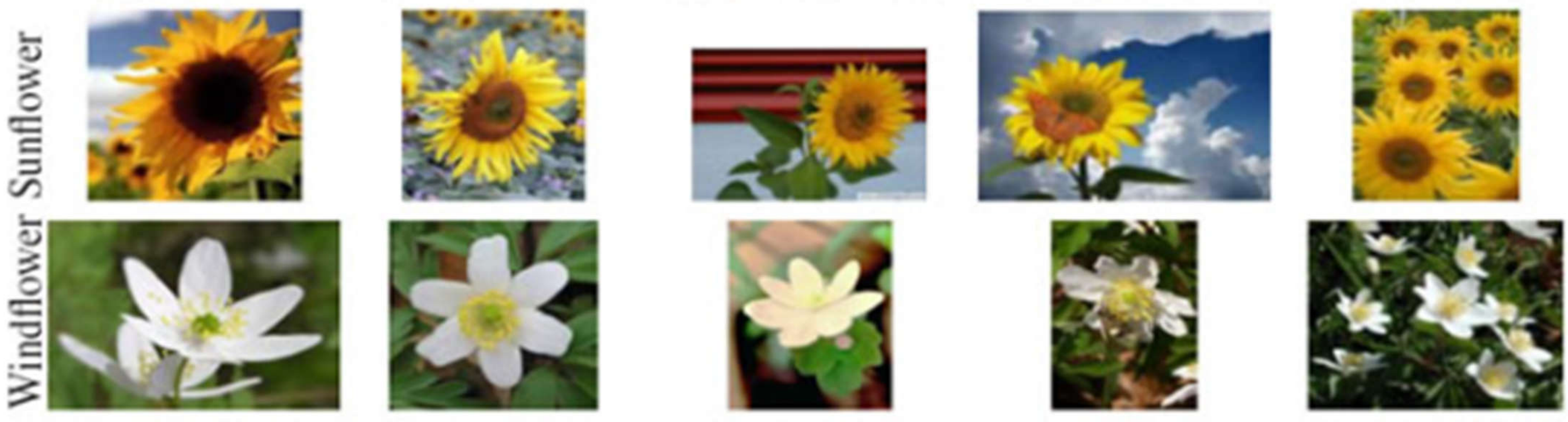


Figure 1. Oxford Flower

- Different objects also usually share common patterns.

Purpose of paper

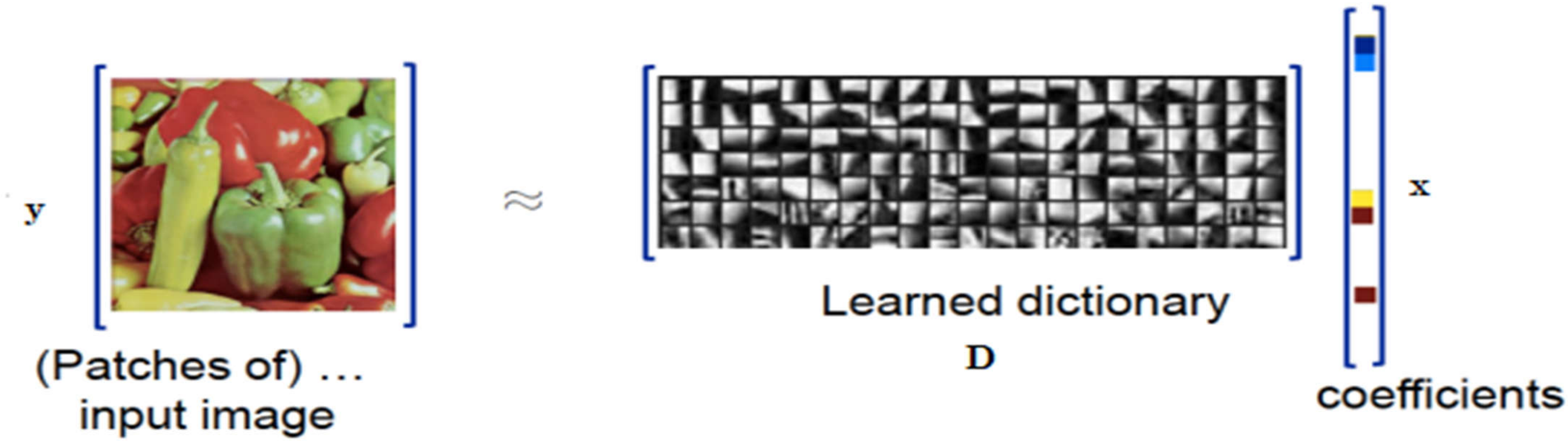
- Build dictionary learning framework is characterized by particular dictionaries and a shared dictionary.

- For the shared dictionary, the authors enforce a low-rank constraint, claim that its spanning subspace has low dimension and the coefficients corresponding to this dictionary is similar.

- For the particular dictionaries, the authors impose on them the well-known constraints stated in the Fisher discrimination dictionary learning.

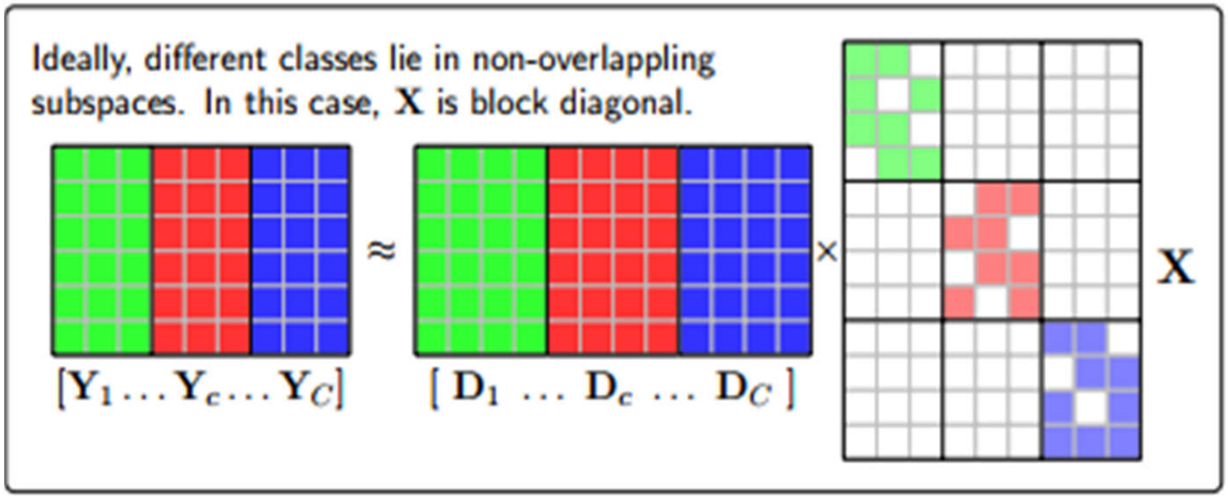
Preparing Knowledge

- Sparse representation for a image



Preparing Knowledge

- Sparse representation for all test image samples – SRC model [2]



- $\mathrm{Y} = [\mathrm{Y}_1, \ldots, \mathrm{Y}_c, \ldots, \mathrm{Y}_C]$ is a set of all test image samples where Y_c comprises those in class c.

- $\mathrm{D} = [\mathrm{D}_1, \ldots, \mathrm{D}_c, \ldots, \mathrm{D}_C]$ is a dictionary where D_c comprises training sample from class c.
- X is sparse matrix. In ideal case, X is a block diagonal.

Preparing Knowledge

- Algorithm SRC [2]:

1. **Input**: a matrix of training samples $\mathrm{D} = [\mathrm{D}_1, \dots, \mathrm{D}_c, \dots, \mathrm{D}_C] \in \mathrm{R}^{m \times n}$, a test sample $\mathrm{y} \in \mathrm{R}^m$.

2. Normalize the columns of D to have unit $\ell_2 - \mathrm{norm}$.

3. Solve the $\ell_1 - \mathrm{minimization}$ [3]:

$$\hat{x}_1 = \mathrm{argmin}_x \|x\|_1 \text{ subject to } y = Dx.$$

4. Compute the residual $\mathrm{r}_c(\mathrm{y}) = \|\mathrm{y} - \mathrm{D}\delta_c(\hat{\mathrm{x}}_1)\|, c = \overline{1, C}$ where $\delta_c(\mathrm{x})$ is a characteristic function from R^n to R^n.

5. **Output:** $\mathrm{Identity}\ (\mathrm{y}) = \mathrm{argmin}_c\ \mathrm{r}_c(\mathrm{y})$.

Preparing Knowledge

♦ Algorithm Fisher discrimination dictionary learning (FDDL) [4]: discriminative dictionary D and the sparse coefficient matrix X are learned based on minimizing the following cost function:

$$\mathrm{J}(\mathrm{D},\mathrm{X}) = \frac{1}{2}\sum_{c=1}^{C} \mathrm{r}(\mathrm{Y_c},\mathrm{D},\mathrm{X_c}) + \lambda_1 \|\mathrm{X}\|_1 + \frac{\lambda_2}{2}\mathrm{f}(\mathrm{X})$$

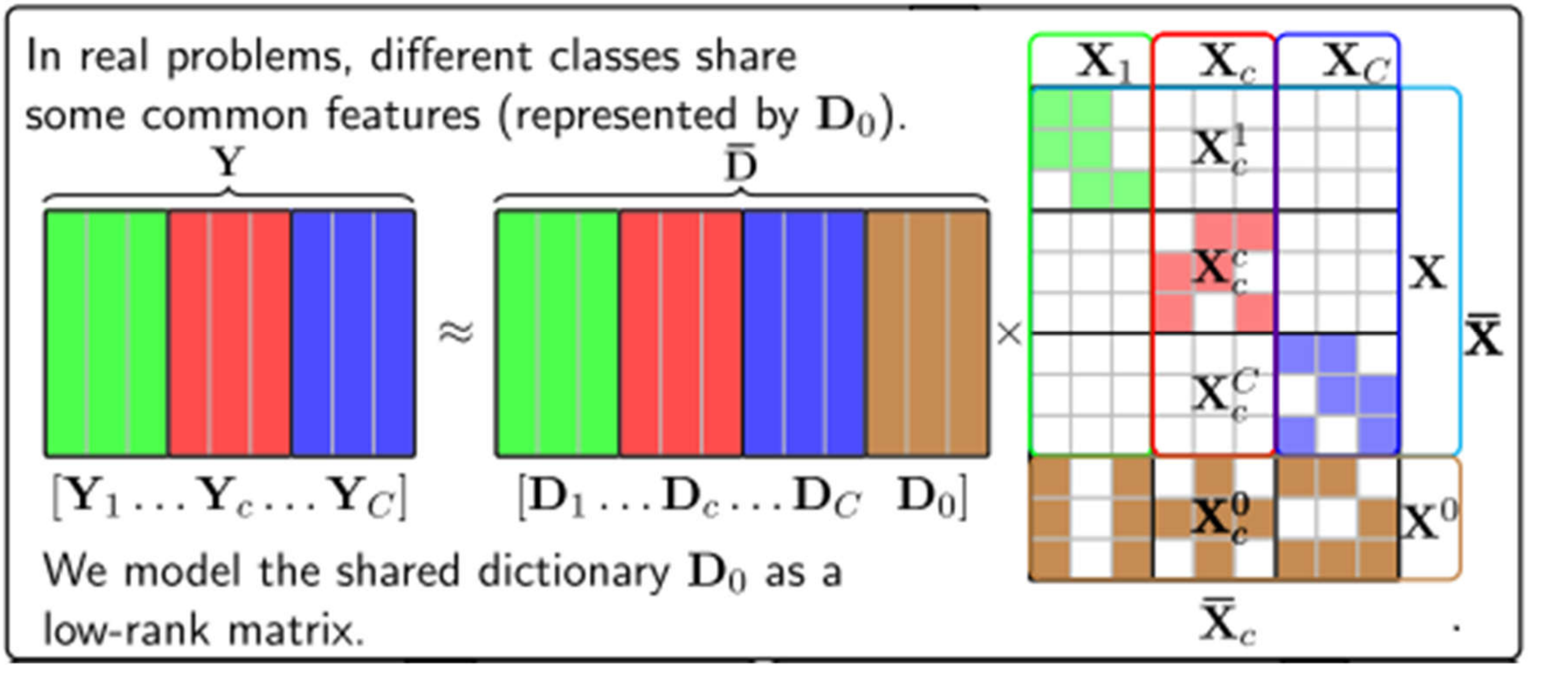
where $\mathrm{r}(\mathrm{Y_c},\mathrm{D},\mathrm{X_c}) = \|\mathrm{Y_c} - \mathrm{DX_c}\|_F^2 + \|\mathrm{Y_c} - \mathrm{DX_c^c}\|_F^2 + \sum_{i\neq c}\|\mathrm{DX_c^i}\|$ and

$$\mathrm{f}(\mathrm{X}) = \sum_{i=1}^{C}\left(\|\mathrm{X_c} - \mathrm{M_c}\|_F^2 - \|\mathrm{M} - \mathrm{M_c}\|_F^2\right) + \|\mathrm{X}\|_F^2$$

with $\mathrm{M_c} = [\mathrm{m_c},\mathrm{m_c},\dots,\mathrm{m_c}]$, $\mathrm{M} = [\mathrm{m},\mathrm{m},\dots,\mathrm{m}]$; $\mathrm{m_c}$ and m are mean vector of $\mathrm{X_c},\mathrm{X}$ columns respectively.

The main idea of paper

- Building dictionary learning framework is characterized by both particular dictionaries and a shared dictionary.



The main idea of paper

- Algorithm Low-Rank Share Dictionary Learning (LRSDL): the dictionary $\overline{\mathrm{D}}$ and the sparse coefficient matrix $\overline{\mathrm{X}}$ are learned based on minimizing the following cost function:

$$\bar{\mathrm{J}}(\overline{\mathrm{D}}, \overline{\mathrm{X}}) = \frac{1}{2}\sum_{c=1}^{C} \bar{\mathrm{r}}(\mathrm{Y_c}, \overline{\mathrm{D}}, \overline{\mathrm{X}}_c) + \lambda_1 \|\overline{\mathrm{X}}\|_1 + \frac{\lambda_2}{2}\bar{\mathrm{f}}(\overline{\mathrm{X}}) + \gamma \|\mathrm{D}_0\|_*$$

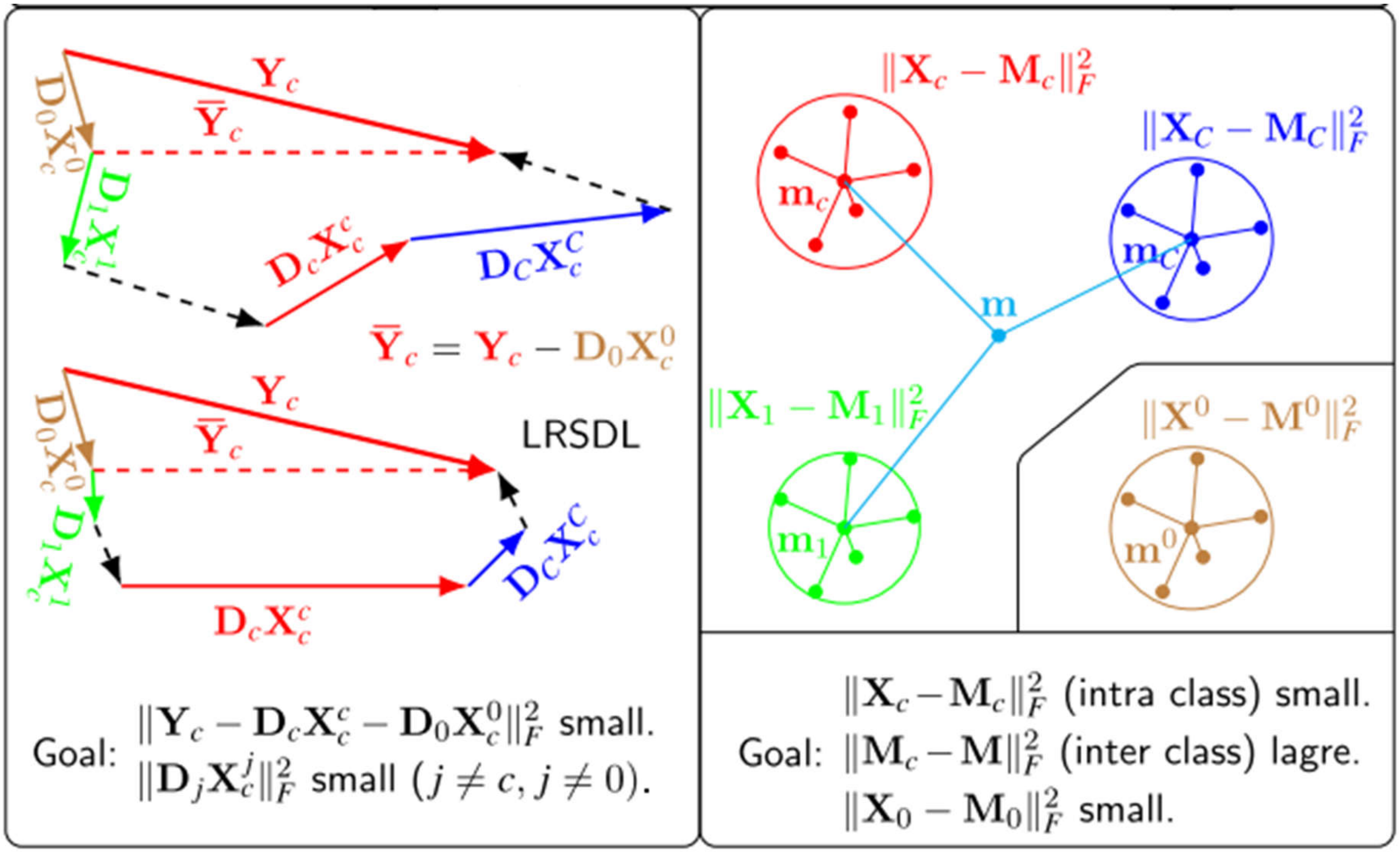
where

$$\bar{\mathrm{r}}(\mathrm{Y_c}, \overline{\mathrm{D}}, \overline{\mathrm{X}}_c) = \|\mathrm{Y_c} - \overline{\mathrm{D}}\overline{\mathrm{X}}_c\|_F^2 + \|\mathrm{Y_c} - \mathrm{D_c X_c^c} - \mathrm{D_0 X_0^c}\|_F^2 + \sum_{j=1, j\neq c}^{C} \|\mathrm{D_j X_c^j}\|_F^2$$

and $\bar{\mathrm{f}}(\overline{\mathrm{X}}) = \mathrm{f}(\mathrm{X}) + \|\mathrm{X}^0 - \mathrm{M}^0\|_F^2$ with M^0 is a mean vector of X^0 columns.

$\|\mathrm{D}_0\|_*$ is a convex relaxion of $\mathrm{rank}(\mathrm{D}_0)$.

The main idea of paper



Algorithm 1 LRSDL Sparse Coefficients Update by FISTA [34]

function $(\hat{\mathbf{X}}, \hat{\mathbf{X}}^0)$ = LRSDL_X$(\mathbf{Y}, \mathbf{D}, \mathbf{D}_0, \mathbf{X}, \mathbf{X}^0, \lambda_1, \lambda_2)$.

1. Calculate:

$$\mathbf{A} = \mathcal{M}(\mathbf{D}^T\mathbf{D}) + 2\lambda_2\mathbf{I};$$
$$\mathbf{B} = 2\mathbf{D}_0^T\mathbf{D}_0 + \lambda_2\mathbf{I}$$
$$L = \lambda_{\max}(\mathbf{A}) + \lambda_{\max}(\mathbf{B}) + 4\lambda_2 + 1^4$$

2. Initialize $\mathbf{W}_1 = \mathbf{Z}_0 = \begin{bmatrix}\mathbf{X}\\ \mathbf{X}^0\end{bmatrix}, t_1 = 1, k = 1$

while not convergence and $k < k_{\max}$ **do**

3. Extract $\mathbf{X}, \mathbf{X}^0$ from $\mathbf{W}_k$.

4. Calculate gradient of two parts:

$$\begin{aligned}
\mathbf{M} &= \mu(\mathbf{X}), \mathbf{M}_c = \mu(\mathbf{X}_c), \widehat{\mathbf{M}} = [\mathbf{M}_1, \dots, \mathbf{M}_C].\\
\mathbf{V} &= \mathbf{Y} - \frac{1}{2}\mathbf{D}\mathcal{M}(\mathbf{X})\\
\mathbf{G} &= \begin{bmatrix}\mathbf{A}\mathbf{X} - \mathcal{M}(\mathbf{D}^T(\mathbf{Y} - \mathbf{D}_0\mathbf{X}^0)) + \lambda_2(\mathbf{M} - \widehat{\mathbf{M}})\\ \mathbf{B}\mathbf{X}^0 - \mathbf{D}_0^T\mathbf{V} - \lambda_2\mu(\mathbf{X}^0)\end{bmatrix}
\end{aligned}$$

5. $\mathbf{Z}_k = \mathcal{S}_{\lambda_1/L}(\mathbf{W}_k - \mathbf{G}/L)$ ($\mathcal{S}_\alpha()$ is the element-wise soft thresholding function. $\mathcal{S}_\alpha(x) = \text{sgn}(x)(|x| - \alpha)_+$).

6. $t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2$

7. $\mathbf{W}_{k+1} = \mathbf{Z}_k + \frac{t_k - 1}{t_{k+1}}(\mathbf{Z}_k - \mathbf{Z}_{k-1})$

8. $k = k + 1$

end while

9. OUTPUT: Extract $\mathbf{X}, \mathbf{X}^0$ from $\mathbf{Z}_k$.

end function

Algorithm 2 LRSDL Algorithm

function $(\hat{\mathbf{X}}, \hat{\mathbf{X}}^0)$ = LRSDL$(\mathbf{Y}, \lambda_1, \lambda_2, \eta)$.

1. Initialization $\mathbf{X} = \mathbf{0}$, and:

$$\begin{aligned}
(\mathbf{D}_c, \mathbf{X}_c^c) &= \arg\min_{\mathbf{D},\mathbf{X}} \frac{1}{2}\|\mathbf{Y}_c - \mathbf{D}\mathbf{X}\|_F^2 + \lambda_1\|\mathbf{X}\|_1\\
(\mathbf{D}_0, \mathbf{X}^0) &= \arg\min_{\mathbf{D},\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \lambda_1\|\mathbf{X}\|_1
\end{aligned}$$

while not converge **do**

2. Update $\mathbf{X}$ and $\mathbf{X}^0$ by Algorithm 1.

3. Update $\mathbf{D}$ by ODL [35]:

$$\begin{aligned}
\mathbf{E} &= (\mathbf{Y} - \mathbf{D}_0\mathbf{X}^0)\mathcal{M}(\mathbf{X}^T)\\
\mathbf{F} &= \mathcal{M}(\mathbf{X}\mathbf{X}^T)\\
\mathbf{D} &= \arg\min_{\mathbf{D}}\{-2\text{trace}(\mathbf{E}\mathbf{D}^T) + \text{trace}(\mathbf{F}\mathbf{D}^T\mathbf{D})\}
\end{aligned}$$

4. Update $\mathbf{D}_0$ by ODL [35] and ADMM [36] (see equations (17) - (20)).

end while

end function

Overall Accuracy (Mean ± Standard Deviation) (%) of Different Dictionary Learning Methods on Different Datasets. Numbers in Parentheses Are Number of Training Samples Per Class

	Ext. YaleB (30)	AR (20)	AR gender (250)	Oxford Flower (60)	Caltech 101 (30)	COIL100 (10)
SRC [5]	97.96 ± 0.22	97.33 ± 0.39	92.57 ± 0.00	75.79 ± 0.23	72.15 ± 0.36	81.45 ± 0.80
LC-KSVD1 [25]	97.09 ± 0.52	97.78 ± 0.36	88.42 ± 1.02	91.47 ± 1.04	73.40 ± 0.64	81.37 ± 0.31
LC-KSVD2 [25]	97.80 ± 0.37	97.70 ± 0.23	90.14 ± 0.45	***92.00 ± 0.73***	73.60 ± 0.53	81.42 ± 0.33
DLSI [29]	96.50 ± 0.85	96.67 ± 1.02	93.86 ± 0.27	85.29 ± 1.12	70.67 ± 0.73	80.67 ± 0.46
DLRD [44]	93.56 ± 1.25	97.83 ± 0.80	92.71 ± 0.43	-	-	-
FDDL [27]	97.52 ± 0.63	96.16 ± 1.16	93.70 ± 0.24	91.17 ± 0.89	72.94 ± 0.26	77.45 ± 1.04
$D^2L^2R^2$ [28]	96.70 ± 0.57	95.33 ± 1.03	93.71 ± 0.87	83.23 ± 1.34	75.26 ± 0.72	76.27 ± 0.98
COPAR [30]	***98.19 ± 0.21***	98.50 ± 0.53	***95.14 ± 0.52***	85.29 ± 0.74	***76.05 ± 0.72***	80.46 ± 0.61
JDL [31]	94.99 ± 0.53	96.00 ± 0.96	93.86 ± 0.43	80.29 ± 0.26	75.90 ± 0.70	80.77 ± 0.85
JDL^* [31]	97.73 ± 0.66	***98.80 ± 0.34***	92.83 ± 0.12	80.29 ± 0.26	73.47 ± 0.67	80.30 ± 1.10
SRRS [45]	97.75 ± 0.58	96.70 ± 1.26	91.28 ± 0.15	88.52 ± 0.64	65.22 + 0.34	**85.04 ± 0.45**
LRSDL	**98.76 ± 0.23**	**98.87 ± 0.43**	**95.42 ± 0.48**	**92.58 ± 0.62**	**76.70 ± 0.42**	***84.35 ± 0.37***

References

[1] T. H. Vu, V. Monga, *Fast Low-Rank Shared Dictionary Learning For Image Classification,* IEEE Trans. Image Process., vol. 26, no. 11, pp. 5160-5175, Nov. 2017.

[2] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, *Robust face recognition via sparse representation*, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[3] S. Chen, D. Donoho, and M. Saunders, *Atomic Decomposition by Basis Pursuit*, SIAM Rev., vol. 43, no. 1, pp. 129-159, 2001.

[4] M. Yang, L. Zhang, X. Feng, and D. Zhang, *Fisher discrimination dictionary learning for sparse representation*, in Proc. IEEE International Conference on Computer Vision, pp. 543–550, Nov. 2011.