STRUCTURED ANALYSIS DICTIONARY LEARNING FOR IMAGE CLASSIFICATION

Wen Tang, Ashkan Panahi, Hamid Krim, Liyi Dai[†] [*]

Department of Electrical and Computer Engerineering
North Carolina State University, Raleigh, NC, USA
[†]Army Research Office, RTP, Raleigh, NC, USA
{wtang6, apanahi, ahk}@ncsu.edu, liyi.dai@us.army.mil

ABSTRACT

We propose a computationally efficient and high-performance classification algorithm by incorporating class structural information in analysis dictionary learning. To achieve more consistent classification, we associate a class characteristic structure of independent subspaces and impose it on the classification error constrained analysis dictionary learning. Experiments demonstrate that our method achieves a comparable or better performance than the state-of-the-art algorithms in a variety of visual classification tasks. In addition, our method greatly reduces the training and testing computational complexity.

Index Terms— Discriminate analysis dictionary learning, structured mapping, supervised learning.

1. INTRODUCTION

Sparse representation has been successfully applied in various image processing and computer vision problems, such as image denoising, and image restoration. Dictionary learning is one way of obtaining sparse representations for signals with unknown precise model. The resulting sparse representation as a linear combination of atoms varies according to the type of dictionary learning techniques: Synthesis Dictionary Learning(SDL) and Analysis Dictionary Learning(ADL).

In contrast to SDL, which assumes that the interesting signal can be recovered by a dictionary with corresponding sparse coefficients, ADL is based on applying the dictionary to the data to yield sparse coefficients.

Due to the success of dictionary learning in image restoration problems, task-driven dictionary learning methods are of great interest in many inference problems, such as image classification. There are broadly two strategies to address the task-driven dictionary learning method. The first strategy is to learn multiple class-specific sub-dictionaries to make the dictionary more structured, and to increase overall discrimination between different classes [1, 2, 3, 4]. To be structured, the atoms in the dictionary are made to learn their own class labels. A class label for a new image can then be decided by comparing reconstruction error from different classes. Another strategy is to learn a shared dictionary for all classes and jointly learn a universal classifier to enforce more discriminative sparse representations [5, 6].

All of the above mentioned techniques have been developed and implemented in the SDL framework, while ADL has increasingly received attention[7]. To the best of our knowledge, none of the standard ADL algorithm such as the analysis K-SVD[8] or the Sparse Null Space(SNS) pursuit [9] has addressed the task driven ADL problem. Shekhar *et al.* [10] have adopted ADL together with SVM to digits and face recognition, and demonstrated that ADL is more stable under noise and occlusion with a competitive performance with SDL. Guo *et al.* [11] integrated local topological structures and discriminative sparse labels into the ADL to yield a k Nearest Neighbor method to classify images.

Inspired by these past efforts and efficient coding of ADL, we propose an integration of structured subspace regularization and supervised learning into an ADL model to obtain a more structured discriminative and efficient approach to image classification. It has been shown, for example in the context of sparse subspace clustering [12], that the sparse representations of the data within a class share a low dimensional subspace. A structuring block diagonal matrix therefore is introduced to achieve these localized subspaces of the sparse codes. This yields more coherence for within-class sparse representations and more disparity for between-class representations. To induce additional robustness in the sought sparse representation, a one-against-all regression-based classifier is jointly learned, with a resulting optimization functional which we solve by a linearized alternating direction method (ADM)[13]. This approach is computationally more efficient than analysis K-SVD[8] and SNS pursuit [9]. Moreover, a great advantage of our algorithm is its extremely short on-line encoding and classification time. Our experiments demonstrate that our method achieves a better overall performances than the synthesis dictionary approach.

The balance of this paper is organized as follows: In Section 2, we state and formulate the problem. We discuss the

[*]Thanks to US-ARO agreement W911NF-16-2-0005.

resulting solution to the optimization problem in Section 3. The experimental validation and results are comprehensively presented in Section 4. We finally provide some concluding comments in Section 5.

2. STRUCTURED ANALYSIS DICTIONARY LEARNING

Notation: Uppercase and lowercase letters respectively denote matrix and vectors. The transpose and inverse of matrix are represented as the superscripts T and -1, such as A^T and A^{-1}. $(a_i)_j$ represents the jth element in the ith column of matrix A.

2.1. ADL Formulation

Given a data matrix $X = [x_1, \ldots, x_n] \in \mathbb{R}^{m \times n}$, the originally formulated ADL[8] problem seeks a representation frame Ω with a sparse coefficient set U.

$$\begin{aligned} &\arg\min_{\Omega, U} \frac{1}{2}\|U - \Omega X\|_2^2 + \lambda_1 \|U\|_1 \\ &s.t.\ \Omega \in \mathbb{R}^{r \times m} \subset \mathcal{W}, \end{aligned} \tag{1}$$

where $U \in \mathbb{R}^{r \times n}$ and $\mathcal{W}$ is a non-trivial solution set.

2.2. Mitigating Inter-Class Feature Interference

The basic idea in our algorithm is to employ the representation U to obtain a classifier. To reduce the impact of inter-class common atoms on the discriminative power of ADL, we propose two additional constraints on U by way of: (1) A structural map of U to minimize interference of inter-class common features. (2) A classification error performance minimization.

***(1) Structural Mapping of U*:** This constraint is particularly enforced by imposing that each class belongs to a subspace defined by a span of the associated coefficients. This improves the consistency of the analysis representations within a class and enhances the divergence between different classes. A block-diagonal matrix $H \in \mathbb{R}^{s \times n}$ as shown below is hence introduced in the training phase,

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

where $s \geq n$ is the length of structured representation. Each diagonal block represents a class and each column h_i^j is a structured representation for the corresponding data point i in the jth class. This constraint may also be deviated by an error term, to be jointly minimized with the ADL functional,

$$H = QU + \varepsilon_1, \tag{2}$$

where $Q \in \mathbb{R}^{s \times r}$ is matrix to be learned with Ω and U, ε_1 is the tolerance.

***(2) Minimal Classification Error*:** The second constraint is a classification error as a feedback term to the learning process of Ω and U. A regression-based classifier $W \in \mathbb{R}^{c \times s}$ is applied to the structured representations QU in this term. We write it as

$$L = W(QU) + \varepsilon_2, \tag{3}$$

where ε_2 is also the tolerance, and the label matrix $L \in \mathbb{R}^{c \times n}$, with c denoting for the number of classes. If image j belongs to class i, $L_{ij} = 1$; otherwise, $L_{ij} = 0$.

2.3. Structured ADL Formulation

To ensure that the structure for each image class is preserved together with minimal interference between different classes, the minimization of tolerance errors is also required. Then, using Eqs.(2), (3) and the minimization of tolerance errors together, the resulting algorithm formulation for our structured ADL is written as

$$\begin{aligned} \arg\min_{\substack{\Omega, U, Q, \\ W, \varepsilon_1, \varepsilon_2}} & \frac{1}{2}\|U - \Omega X\|_F^2 + \lambda_1 \|U\|_1 \\ & + \frac{\rho_1}{2}\|\varepsilon_1\|_2^2 + \frac{\rho_2}{2}\|\varepsilon_2\|_2^2 \\ s.t.\ & H = QU + \varepsilon_1, \\ & L = W(QU) + \varepsilon_2, \\ & \|\omega_i^T\|_2^2 = 1; \forall i = 1, \ldots, r, \end{aligned} \tag{4}$$

where ω_i^T is the row of Ω, ρ_1 and ρ_2 are the penalty coefficients. Recall H is the structured representation, Q is the structuring transformation, L is the classifier label, and W is the linear classifier, and λ_1 is the tuning parameters.

3. ALGORITHMIC SOLUTION

The objective function in Eq.(4), on account of its non-convexity, is transformed to an augmented Lagrange formulation with dual variables $Y^{(1)}$, $Y^{(2)}$ and μ. After straight forward calculations that lead to eliminations of ε_1 and ε_2, we obtain the following expression for this function:

$$\begin{aligned} &L(\Omega, U, Q, W, Y^{(1)}, Y^{(2)}, \mu) = \frac{1}{2}\|U - \Omega X\|_F^2 + \lambda_1 \|U\|_1 \\ &+ \lambda_2 < Y^{(1)}, H - QU > + \lambda_3 < Y^{(2)}, L - W(QU) > \\ &+ \frac{\mu}{2}\|H - QU\|_2^2 + \frac{\mu}{2}\|L - W(QU)\|_2^2, \end{aligned} \tag{5}$$

where $\lambda_1, \lambda_2, \lambda_3 > 0$ are the new tuning parameters. Then, to minimize the objective functional in Eq.(5), we first randomly initialize the analysis dictionary Ω and two linear transformations Q and W. The sparse representation U is initialized by $U = \mathbf{0}$, the zero matrix. η_Q, η_{WQ}, and $\eta_{WU} > 0$ are the

parameters for the learning rate. Then, we alternately update different variables when fixing the others, which is summarized in Algorithm 1.

Algorithm 1 Structured Analysis Dictionary Learning

Initialize Ω, Q, and W as random matrices, and initialize U as a zero matrix; T is maximum iteration;
while not converged **and** $k < T$ **do**
  $k = k + 1$;
  $U_{k+1} = \tau_{\frac{\lambda_1}{\mu_k(\eta_Q+\eta_{WQ})}}\left(U_k - \frac{\nabla_U L(\Omega_k,U_k,Q_k,W_k,Y_k^{(1)},Y_k^{(2)})}{\mu_k(\eta_Q+\eta_{WQ})}\right)$;
  $Q_{k+1} = Q_k - \frac{\nabla_Q L(\Omega_k,U_{k+1},Q_k,W_k,Y_k^{(1)},Y_k^{(2)})}{\mu_k(\eta_Q+\eta_{WU})}$,
  $W_{k+1} = W_k - \frac{\nabla_W L(\Omega_k,U_{k+1},Q_{k+1},W_k,Y_k^{(1)},Y_k^{(2)})}{\mu_k\eta_{QU}}$;
  $\Omega_{k+1} = U_{k+1}X^T(XX^T + \lambda_4 I)^{-1}$;
  Normalize Ω_{k+1} by $\omega_i^T = \frac{\omega_i^T}{\|\omega_i^T\|_2}, \forall i$;
  $Y_{k+1}^{(1)} = Y_k^{(1)} + \mu_k(H - Q_{k+1}U_{k+1})$;
  $Y_{k+1}^{(2)} = Y_k^{(2)} + \mu_k(L - W_{k+1}Q_{k+1}U_{k+1})$;
  $\mu_{k+1} = \min\{\rho\mu_k, \mu_{max}\}$; $\%\rho$ is the learning rate
end while

4. EXPERIMENTS AND RESULTS

We evaluate our proposed SADL method on four popular visual classification datasets which have been widely used in previous works and with known performance benchmarks. They include Extended YaleB[14] face dataset, AR[15] face dataset, Caltech101[16] object categorization dataset and Scene15[17] scene image dataset. The features of these 4 datasets are extracted by the same settings in [6].

In our experiments, we provide a comparative evaluation of three state-of-the-art techniques and our proposed technique, including classification accuracy and training and testing times. The testing time is defined as the average processing time to classify a single image. For a fair comparison, we measure the performances of all algorithms by using the same dictionary size on each dataset and experiment over 10 realizations to obtain an average performance. In relation to competitive methods, ADL+SVM [10] is a baseline. SRC [1] is the classical Sparse Representation based Classification. LC-KSVD [6] is a SDL approach that jointly learns a discriminative dictionary and a universal classifier. In our tables, the accuracy in the parentheses with the citation is the one that was reported in the original paper. The difference of the accuracy of our implementing and the original one might be caused by the different segmentations of the training and testing samples.

4.1. Face Recognition

***Extended YaleB*:** This face dataset contains in total 2414 frontal face images of 38 persons under various illumination and expression conditions, as illustrated in Fig.1. Each Extended YaleB face image has a 504-dimensional feature vector. We randomly choose half of the images for training, and the rest for testing. The dictionary size is set to 570 atoms, $\lambda_1 = 0.001$, $\lambda_2 = 9$, $\lambda_3 = 3$, $\lambda_4 = 0.5$ and $T = 780$.

Fig. 1. Examples of Face Dataset: The left figure is Extended YaleB Dataset, and the right one is AR Dataset.

Table 1. Classification Results on Extended YaleB Dataset

Methods	Accuracy (%)	Training (s)	Testing (s)
ADL+SVM[10]	82.91%	91.78	1.13×10^{-3}
SRC[1]	80.5%	No Need	3.74×10^{-1}
LC-KSVD[6]	94.56% (**95**% [6])	234.67	1.63×10^{-2}
SADL	94.91%	**51.29**	**2.72**$\times10^{-6}$

The classification results, training and testing times are summarized in Table 1. Our proposed SADL method achieves the highest classification accuracy in the test, but tinily lower than the reported accuracy of LC-KSVD. However, it is still substantially more efficient than the others in terms of numerical complexity and classification .

For a more thorough evaluation, we compare SADL with LC-KSVD for different dictionary sizes, and display the classification accuracy in Fig.2. We ran our experiments for dictionary sizes by 32, 128, 224, 320, 416, 512, 608, 704, 800, 896, 992, and 1216 (all training size). SADL exhibits a more stable performance than that of LC-KSVD. In particular, the accuracy of LC-KSVD significantly decreases, when the dictionary size approaches the all training sample size. In addition, our method apparently has a much higher classification accuracy than LC-KSVD, when the dictionary size is small. The significant decrease in accuracy may be caused by the trivial solution of dictionary in SDL.

***AR*:** The AR face dateset has 2600 color images of 50 females and 50 males with more facial variations than the Extended YaleB database, such as different illumination conditions, expressions and facial disguises, as shown in Fig. 1. Each person has about 26 images of size 165×120. The AR Face feature dimension is 540. 20 images of each person are randomly selected as a training set and the other 6 images for testing. The dictionary size of the AR dataset is set to 500 atoms, $\lambda_1 = 0.001$, $\lambda_2 = 8$, $\lambda_3 = 10$, $\lambda_4 = 0.5$ and $T = 1040$.

The classification performances are summarized in Table 2. Our proposed SADL achieves higher classification accuracy than others. Our method is about 10000 times faster than SRC and LC-KSVD for the testing phase.

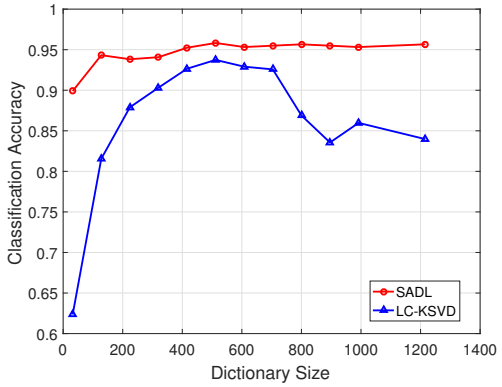

Fig. 2. Classification Accuracy

Table 2. Classification Results on AR Dataset

Methods	Accuracy (%)	Training (s)	Testing (s)
ADL+SVM[10]	90.40%	218.54	9.10×10^{-3}
SRC[1]	66.50%	No Need	5.25×10^{-2}
LC-KSVD[6]	87.78% (93.7%[6])	244.52	1.42×10^{-2}
SADL	**95.08%**	**89.13**	**3.67**$\times10^{-6}$

4.2. Object Recognition

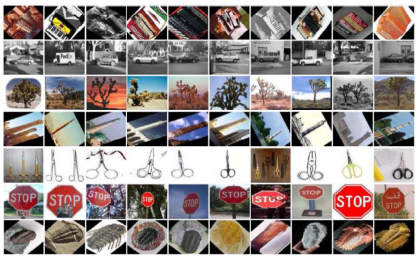

Fig. 3. Caltech101 Dataset Examples

The Caltech101 dataset has 101 different categories of different objects and 1 non-object category. Most categories have around 50 images. Fig.3 gives some examples from the Caltech101 dataset. The standard bag-of words+spatial pyramid matching (SPM) framework [17] is used to calculate the SPM features. PCA is then adopted to reduce the dimension of a SPM feature to 3000. The dictionary size is set to 510, $\lambda_1 = 0.001$, $\lambda_2 = 10$, $\lambda_3 = 3$, $\lambda_4 = 4.6$ and $T = 990$.

We evaluate all methods with a dictionary size of 510. The classification performances are summarized in Table 3. Our proposed SADL still achieves the highest performance of the lot. SADL has again a short testing time, which is around 10000 times faster than LC-KSVD.

4.3. Scene Classification

Scene15 dataset contains a total of 15 categories of different scenes, and each category has around 200 images. The examples are listed in Fig.4. Proceeding as for the Caltech 101 dataset, we compute the SPM features for scene images. Each scene image is transformed to a 3000 dimensional feature by PCA. We randomly pick 100 images per class as training data, and use the rest of images as testing data. The settings and steps follow [6]. The dictionary size is set to 450, $\lambda_1 = 0.001$, $\lambda_2 = 10$, $\lambda_3 = 4$, $\lambda_4 = 0.001$ and $T = 220$.

Table 3. Classification Results on Caltech101 Dataset

Methods	Accuracy (%)	Training (s)	Testing (s)
ADL+SVM[10]	54.93%	**447.80**	7.75×10^{-3}
SRC[1]	67.70%	No Need	4.34×10^{-1}
LC-KSVD[6]	71.79%	487.61	1.35×10^{-2}
SADL	**72.36%**	773.66	**8.10**$\times10^{-6}$

Fig. 4. Scene15 Dataset Examples

Table 4. Classification Results on Scene15 Dataset

Methods	Accuracy (%)	Training (s)	Testing (s)
ADL+SVM[10]	49.35%	**110.47**	1.14×10^{-4}
SRC[1]	91.80%	No Need	4.06×10^{-1}
LC-KSVD[6]	**98.83%** (92.9%[6])	270.93	1.26×10^{-2}
SADL	98.16%	121.02	**9.23**$\times10^{-6}$

The classification performances are summarized in Table 4. Our performance is slightly lower than LC-KSVD, but is still higher than SRC, ADL+SVM and the LC-KSVD reported accuracy. However, the testing phase is superior to the others. Note that, the testing time is 10 thousand times faster than LC-KSVD.

5. CONCLUSION

We proposed an image classification method referred to as structured analysis dictionary learning (SADL). To obtain SADL, we constrained a structured subspace(cluster) model in the enhanced ADL method, where each class was represented by a structured subspace. The enhancement of ADL was realized by constraining the learning by a classification fidelity term on the sparse coefficients. Our formulated optimization problem was efficiently solved by the linearized ADM method, in spite of its non-convexity due to bilinearity. Taking advantage of analysis dictionary, our method achieved a significantly faster testing time.

6. REFERENCES

[1] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma, "Robust face recognition via sparse representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[2] Ignacio Ramirez, Pablo Sprechmann, and Guillermo Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 3501–3508.

[3] Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang, "Fisher discrimination dictionary learning for sparse representation," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 543–550.

[4] Zhaowen Wang, Jianchao Yang, Nasser Nasrabadi, and Thomas Huang, "A max-margin perspective on sparse representation-based classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1217–1224.

[5] Julien Mairal, Jean Ponce, Guillermo Sapiro, Andrew Zisserman, and Francis R Bach, "Supervised dictionary learning," in *Advances in neural information processing systems*, 2009, pp. 1033–1040.

[6] Zhuolin Jiang, Zhe Lin, and Larry S Davis, "Label consistent k-svd: Learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[7] Sangnam Nam, Mike E Davies, Michael Elad, and Rémi Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[8] Ron Rubinstein, Tomer Peleg, and Michael Elad, "Analysis k-svd: A dictionary-learning algorithm for the analysis sparse model," *Signal Processing, IEEE Transactions on*, vol. 61, no. 3, pp. 661–677, 2013.

[9] Xiao Bian, Hamid Krim, Alex Bronstein, and Liyi Dai, "Sparsity and nullity: Paradigms for analysis dictionary learning," *SIAM Journal on Imaging Sciences*, vol. 9, no. 3, pp. 1107–1126, 2016.

[10] Sumit Shekhar, Vishal M Patel, and Rama Chellappa, "Analysis sparse coding models for image-based classification," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 5207–5211.

[11] Jun Guo, Yanqing Guo, Xiangwei Kong, Man Zhang, and Ran He, "Discriminative analysis dictionary learning," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[12] Ehsan Elhamifar and Rene Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

[13] Zhouchen Lin, Risheng Liu, and Zhixun Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Advances in neural information processing systems*, 2011, pp. 612–620.

[14] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[15] A.M. Martinez and R. Benavente, "The ar face database," *CVC Technical Report*, , no. 24, June 1998.

[16] Li Fei-Fei, Rob Fergus, and Pietro Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 59–70, 2007.

[17] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. IEEE, 2006, vol. 2, pp. 2169–2178.

[18] Wen Tang, Ives Rey Otero, Hamid Krim, and Liyi Dai, "Analysis dictionary learning for scene classification," in *Statistical Signal Processing Workshop (SSP), 2016 IEEE*. IEEE, 2016, pp. 1–5.

[19] Shahin Mahdizadehaghdam, Liyi Dai, Hamid Krim, Erik Skau, and Han Wang, "Image classification: A hierarchical dictionary learning approach," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2597–2601.