STRUCTURE-CONSTRAINED LOW-RANK AND PARTIAL SPARSE REPRESENTATION FOR IMAGE CLASSIFICATION

Yang LIU[⋆], *Haixu LIU*[⋆], *Chenyu LIU*[⋆], *Xueming LI*[⋆†],

[⋆]School of Information and Communication Engineering
Beijing University of Posts and Telecommunications, Beijing, China 100876
[†]Beijing Key Laboratory of Network System and Network Culture

ABSTRACT

In this paper, a novel Structure-Constrained Low-Rank and Partial Sparse Representation algorithm for image classification is proposed. First, a Structure-Constrained Low-Rank dictionary learning algorithm is proposed, which imposes both structure and low-rank restriction on the coefficient matrix. Second, under the assumption that the representation of test sample is sparse and correlated with the learned representation of training samples, we concatenate training samples and test samples to form a data matrix and find a low-rank and sparse representation of the data matrix over learned dictionary by low-rank matrix recovery technique. Experimental results demonstrate the effectiveness of the proposed algorithm.

Index Terms**—** sparse representation, low-rank representation, dictionary learning, image classification

1. INTRODUCTION

The recent studies show that sparsity is a ubiquitous property exhibited by many real-world signals such as audio and images, and sparsity also acts as a strong prior for solving the ill-posed inverse problems [1]. Thus sparse representation, has drawn considerable interest in recent years [2] and is extensively used in image processing [3,4]. The problem solved by sparse representation is to approximate the given signal by sparse linear combinations of elements on a basis or an over-complete dictionary. Sparse representation achieves inspiring performance on image classification. Sparse representation based classification (SRC) [5] takes all training samples as dictionary. However, to obtain a high recognition accuracy, the size of dictionary should be large while the optimization is computationally expensive. Some dictionary learning algorithm is therefore proposed to learn a compact and discriminative dictionary [6–8].

Low rank matrix recovery, which determines a low-rank matrix from given data matrix, has received an increasing amount of interest in recent years. It has been successfully applied to a variety of applications, such as image classification [9–12], background modeling from surveillance video [13], matrix completion [14] and subspace clustering [15].

Recently some image classification algorithms based on low-rank and sparse representation have been proposed. Ma [9] proposed a discriminative low-rank dictionary learning for sparse representation (DLRD-SR), which learns low-rank class-specific sub-dictionary of every class and combines all sub-dictionaries to form a dictionary. However, the low-rankness of each class-specific sub-dictionary leads to a decrease of the representative power of the dictionary. Zhang [10] proposed a structured low-rank representation for image classification, which assumes that the coefficient matrix should be both low-rank and sparse, Zhang also constructed an structured ideal representation based on the label information of training samples and encouraged the coefficient matrix to be close to the ideal representation. However, the magnitude of the ideal representation is hard to approximate and this method can only find low-rank and sparse representation of test samples after all test samples are received, which is not a practical way.

In this paper, a novel Structure-Constrained Low-Rank and Partial Sparse Representation (SC-LCPSR) algorithm, which is consisted of a Structure-Constrained Low-Rank Dictionary Learning algorithm and a Low-Rank and Partial Sparse Representation algorithm, for image classification is proposed. First, SC-LCPSR learns a structured discriminative dictionary by low-rank and structure constraint on the coefficient matrix and then a data matrix is formed by combining test sample and training samples, and the low-rank and sparse representation of the data matrix under the learned dictionary is found on the assumption that the representation of test sample is sparse and correlated with the learned representation of training samples. Further, the proposed algorithm handles one test sample every time, which is a more practical way, and it is also applicable to the situation that multiple test samples are received at the same time.

The rest of this paper is organized as follows. Some related work is reviewed in Section 2. The proposed algorithm is introduced in Section 3. Experimental results is demonstrated in Section 4. Finally Section 5 concludes this paper.

2. RELATED WORK

In most image processing and computer vision applications, the given data matrix $\mathbf{X}$ can be decomposed as a low-rank component and a sparse component. Candès et.al proposed Robust PCA [13] to model this kind of data:

$$\arg\min_{\mathbf{Z},\mathbf{E}} \ \text{rank}(\mathbf{Z}) + \lambda\|\mathbf{E}\|_0 \quad \text{s.t.}\ \mathbf{X} = \mathbf{Z} + \mathbf{E} \tag{1}$$

However, to find the solution of (1) is NP-hard. It is proved in [13] that if the rank of $\mathbf{Z}$ is low enough and $\mathbf{E}$ is sparse enough, (1) can be relaxed into a convex problem:

$$\arg\min_{\mathbf{Z},\mathbf{E}} \ \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_1 \quad \text{s.t.}\ \mathbf{X} = \mathbf{Z} + \mathbf{E} \tag{2}$$

where $\|\mathbf{Z}\|_*$ is the nuclear norm of $\mathbf{Z}$ (i.e. the sum of singular value).

This model implicitly assumes that the underlying data structure is a single low-rank subspace. When the data is drawn from a union of multiple subspaces, which is common in image classification, the recovery may be inaccurate. Lin [16] proposed a more general rank minimization problem Low-Rank Representation (LRR) to solve this problem, and LRR is defined as follows:

$$\arg\min_{\mathbf{Z},\mathbf{E}} \ \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_1 \quad \text{s.t.}\ \mathbf{X} = \mathbf{D}\mathbf{Z} + \mathbf{E} \tag{3}$$

where $\mathbf{D}$ is a dictionary that linearly spans the data space.

Sparse Representation (SR) is also a powerful tool and has attracted significant research interest recently. But as observed in [17], SR focus on local structure of data for the sparsest representation of each data vector is found individually and no global constraint on it, while LRR is better at capturing the global structure of the data matrix.

Let's considerate a simple example.

$$\arg\min_{\mathbf{Z}} \ R(\mathbf{Z}) \quad \mathbf{X} = \mathbf{X}\mathbf{Z} \tag{4}$$

where $R(\mathbf{Z})$ denotes low-rank or sparse constraint on $\mathbf{Z}$. When we solve problem (4) via SR directly, a trial solution $\mathbf{Z} = \mathbf{I}$ is found, which is the most sparse solution but it is useless, an additional condition $\text{diag}(\mathbf{Z}) = \mathbf{0}$ is needed to find the solution we pursuit [15], while the optimization problem (4) can be solved under low-rank constraint directly [16]. This characteristic of LLR is extensively used in subspace segmentation. Image classification can also benefit from it, especially taking the sparsity of the coefficient matrix into consideration at the same time.

3. STRUCTURE-CONSTRAINED LOW-RANK AND PARTIAL SPARSE REPRESENTATION

In this section, we will introduce the proposed Structure-Constrained Low-Rank and Partial Sparse Representation (SC-LCPSR). The proposed SC-LCPSR is consisted of two parts, a Structure-Constrained Low-Rank Dictionary Learning algorithm and a Low-Rank and Partial Sparse Representation algorithm.

3.1. Structure-Constrained Low-Rank Dictionary Learning

Let $\mathbf{X}^t = [\mathbf{X}_1^t, \mathbf{X}_2^t, \ldots, \mathbf{X}_{Nc}^t] \in \mathbb{R}^{p\times Ns}$, where $\mathbf{X}_i^t$ denotes training samples from the i^{th} class, Nc means number of classes, Ns means number of samples and $Ns = \sum_{i=1}^{Nc} N_i$, where N_i denotes the number of samples from the i^{th} class. Low-rank representation decomposes the matrix $\mathbf{X}_i^t$ into a low-rank component $\mathbf{D}\mathbf{Z}_i^t$ and noise $\mathbf{E}_i$, $i.e.$, $\mathbf{X}_i^t = \mathbf{D}\mathbf{Z}_i^t + \mathbf{E}_i$, with respect to a dictionary $\mathbf{D}$. It's assumed that every atom of $\mathbf{D}$ is associated with a class, or every class contributes several atoms in $\mathbf{D}$, $i.e.$, $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_{Nc}]$. Consequently $\mathbf{X}_i^t$ tends to be represented by atoms associated with the i^{th} class.

Based on the above discussion, it's possible to learn a discriminative dictionary $\mathbf{D}$, a low-rank representation $\mathbf{Z}_i^t$ and noise $\mathbf{E}$ with given training samples from the i^{th} class $\mathbf{X}_i^t$ under some constrain on $\mathbf{Z}_i^t$. The objective function is formulated as follows:

$$\begin{array}{ll} \arg\min\limits_{\mathbf{Z}_i^t,\mathbf{E}_i,\mathbf{D}} & \sum_{i=1}^{Nc} \alpha\|\mathbf{Z}_i^t\|_* + \sum_{i=1}^{Nc} \gamma\|\mathbf{E}_i\|_l \\ \text{s.t.} & \mathbf{X}_i^t = \mathbf{D}\mathbf{Z}_i^t + \mathbf{E}_i \\ & \pi_{\Omega_i}(\mathbf{Z}_i^t) = \mathbf{0} \end{array} \tag{5}$$

where $\|\mathbf{E}_i\|_l$ indicates certain regularization strategy, such as $\|\mathbf{E}_i\|_F^2$ for the data is slightly perturbed, $\|\mathbf{E}_i\|_1$ for random corruptions and $\|\mathbf{E}_i\|_{2,1}$ for sample-specific corruptions, here, $\|\mathbf{E}\|_l$ is set to be $\|\mathbf{E}\|_1$; α and γ are parameters that balance the two term; Ω_i denotes the rows in $\mathbf{Z}_i^t$ corresponding to D_i, $\pi_{\Omega_i} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is a linear operator that keeps the entry in Ω_i unchanged and sets those outside Ω_i zeros. The constraint $\pi_{\Omega_i}(\mathbf{Z}_i^t) = \mathbf{0}$ incorporates the label information in training samples and also makes $\mathbf{Z}_i^t$ sparse.

The optimization of (5) can be divided into two subproblem. The first subproblem is to compute $\mathbf{Z}_i^t$ and $\mathbf{E}_i$ with $\mathbf{D}$ fixed, while the second subproblem is to solve $\mathbf{D}$ with given $\mathbf{Z}_i^t$ and $\mathbf{E}_i$.

The solution of $\mathbf{Z}_i^t$ and $\mathbf{E}_i$ can be found by alternatively solving each pair of $\mathbf{Z}_i^t$ and $\mathbf{E}_i$ via Inexact Augmented Lagrange Multiplier(ALM) [18] in partial augmented Lagrangian function below:

$$\begin{array}{ll} \arg\min\limits_{\mathbf{J}_i,\mathbf{Z}_i^t,\mathbf{E}_i} & \alpha\|\mathbf{J}_i\|_* + \gamma\|\mathbf{E}_i\|_1 + < \mathbf{Y}_2, \mathbf{Z}_i^t - \mathbf{J}_i > \\ & + < \mathbf{Y}_1, \mathbf{X}_i^t - \mathbf{D}\mathbf{Z}_i^t - \mathbf{E}_i > + \\ & \frac{\mu}{2}(\|\mathbf{X}_i^t - \mathbf{D}\mathbf{Z}_i^t - \mathbf{E}_i\|_F^2 + \|\mathbf{Z}_i^t - \mathbf{J}_i\|_F^2) \end{array} \tag{6}$$

where for updating $\mathbf{Z}_i^t$ the constraint $\pi_{\Omega_i}(\mathbf{Z}_i^t) = \mathbf{0}$ should be enforced after optimization.

$\mathbf{Z}^t$ and $\mathbf{E}$ are formed by concatenating all $\mathbf{Z}_i^t$ and $\mathbf{E}_i$ together respectively, i.e., $\mathbf{Z}^t = [\mathbf{Z}_1^t, \mathbf{Z}_2^t, \ldots, \mathbf{Z}_{Nc}^t]$ and $\mathbf{E} = [\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_{Nc}]$. And the dictionary is updated as follows:

$$\mathbf{D} := \theta\mathbf{D} + (1-\theta)(\mathbf{X}^t - \mathbf{E})(\mathbf{Z}^t)^\dagger \tag{7}$$

where $(\mathbf{Z}^t)^\dagger$ denotes the pseudo-inverse of $\mathbf{Z}^t$, $0 \le \theta \le 1$ is a parameter controls the updating step.

The calculation is described in Algorithm 1. Where $\mathcal{S}_{\alpha\mu^{-1}}[\mathbf{S}]$ means soft-thresholding on $\mathbf{S}$.

Algorithm 1 Structure-Constrained Low-Rank Dictionary Learning Algorithm

Input:
Initialized dictionary $\mathbf{D}$; Dictionary updating parameter θ; Parameters for Inexact ALM

Output:
$\mathbf{D}$, $\mathbf{Z}^t$, $\mathbf{E}$

while not converged **do**
  for $i = 1 \to N_c$ **do**
    $\mathbf{J}_i$=$\mathbf{0}$, $\mathbf{Z}_i^t$=$\mathbf{0}$, $\mathbf{E}_i$=$\mathbf{0}$, $\mathbf{Y}_1$=$\mathbf{0}$, $\mathbf{Y}_2$=$\mathbf{0}$
    while not converged **do**
      $(\mathbf{U}, \mathbf{S}, \mathbf{V}) = \mathrm{svd}(\mathbf{Z}_i^t + \mathbf{Y}_2/\mu)$
      $\mathbf{J}_i := \mathbf{U}\mathcal{S}_{\alpha\mu^{-1}}[\mathbf{S}]\mathbf{V}^T$
      $\mathbf{Z}_i^t := (\mathbf{D}^T\mathbf{D}+\mathbf{I})^{-1}(\mathbf{D}^T\mathbf{X}_i^t - \mathbf{D}^T\mathbf{E}_i + \mathbf{J}_i + (\mathbf{D}^T\mathbf{Y}_1 - \mathbf{Y}_2)/\mu)$
      $\pi_{\Omega_i}(\mathbf{Z}_i^t) = \mathbf{0}$
      $\mathbf{E}_i := \mathcal{S}_{\gamma\mu^{-1}}(\mathbf{X}_i^t - \mathbf{D}\mathbf{Z}_i^t + \mathbf{Y}_1/\mu))$
      $\mathbf{Y}_1 := \mathbf{Y}_1 + \mu(\mathbf{X}_i^t - \mathbf{D}\mathbf{Z}_i^t - \mathbf{E}_i)$
      $\mathbf{Y}_2 := \mathbf{Y}_2 + \mu(\mathbf{Z}_i^t - \mathbf{J}_i)$
      $\mu = \min(\mu_{max}, \rho\mu)$
    end while
  end for
  $\mathbf{Z}^t = [\mathbf{Z}_1^t, \mathbf{Z}_2^t, \ldots, \mathbf{Z}_{Nc}^t]$
  $\mathbf{E} = [\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_{Nc}]$
  $\mathbf{D} := \theta\mathbf{D} + (1-\theta)(\mathbf{X}^t - \mathbf{E})(\mathbf{Z}^t)^\dagger$
end while

3.2. Low-Rank and Partial Sparse Representation

There are several methods to obtain coefficient $\mathbf{z}$, given a test sample $\mathbf{x}$ and learned dictionary $\mathbf{D}$, such as sparse representation. But as mentioned above, the sparsity criterion captures only the local structure of $\mathbf{x}$. The low-rankness captures the global structure of data, but test sample is given one by one in most cases, hence it is inappropriate to wait for the test samples to form a matrix. In this subsection, we proposed a Low-Rank and Partial Sparse Representation algorithm(LRPSR) to capture not only the local structure of $\mathbf{x}$ but also the global structure of $\mathbf{x}$ and $\mathbf{X}^t$. This algorithm process one test sample every time which is a more practical way, and actually this algorithm also applies to the situation that multiple test samples are received at the same time.

It is assumed that the test sample from i^{th} class is highly correlated with the training samples that also come from i^{th} class $\mathbf{X}_i^t$, hence the coefficient $\mathbf{z}$ should be correlated with $\mathbf{Z}_i^t$, but i is unknown. Consequently, we consider that $\mathbf{z}$ is correlated with $\mathbf{Z}^t$ and the correlation leads to low-rankness.

The test sample $\mathbf{x}$ and training samples $\mathbf{X}^t$ are concatenated to form a matrix $\mathbf{X}$, i.e., $\mathbf{X} = [\mathbf{X}^t, \mathbf{x}]$. $\mathbf{X}$ can be decomposed into low-rank component $\mathbf{DZ}$ and noise $\mathbf{E}$, i.e., $\mathbf{X} = \mathbf{DZ} + \mathbf{E}$ and $\mathbf{Z}$ can be divided into two parts, $\mathbf{Z} = \mathbf{F} + \mathbf{W}$, where $\mathbf{F}$ means the fixed part and $\mathbf{F} = [\mathbf{Z}^t, \mathbf{0}]$, $\mathbf{W}$ denotes the coefficient of $\mathbf{x}$ and $\mathbf{W} = [\mathbf{0}, \mathbf{z}]$. By adding low-rank and sparse constrain, the objective function is formulated as follows:

$$\begin{aligned} \arg\min_{\mathbf{Z},\mathbf{W},\mathbf{E}} \quad & \alpha\|\mathbf{Z}\|_* + \beta\|\mathbf{W}\|_1 + \gamma\|\mathbf{E}\|_1 \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{DZ} + \mathbf{E} \\ & \mathbf{Z} = \mathbf{F} + \mathbf{W} \\ & \pi_\Omega(\mathbf{W}) = \mathbf{0} \end{aligned} \tag{8}$$

where $\pi_\Omega(\mathbf{W}) = \mathbf{0}$ sets all entries in $\mathbf{W}$ to zero except columns corresponding to test samples.

The problem (8) can be optimized via Inexact ALM when it is rewritten as partial augmented Lagrangian function:

$$\begin{aligned} \arg\min_{\mathbf{J},\mathbf{Z},\mathbf{W},\mathbf{E}} \quad & \alpha\|\mathbf{J}\|_* + \beta\|\mathbf{W}\|_1 + \gamma\|\mathbf{E}\|_1 + \\ & < \mathbf{Y}_1, \mathbf{X} - \mathbf{DZ} - \mathbf{E} > + < \mathbf{Y}_2, \mathbf{Z} - \mathbf{J} > \\ & + < \mathbf{Y}_3, \mathbf{Z} - \mathbf{F} - \mathbf{W} > + \frac{\mu}{2}(\|\mathbf{Z} - \mathbf{J}\|_F^2 \\ & + \|\mathbf{Z} - \mathbf{F} - \mathbf{W}\|_F^2 + \|\mathbf{X} - \mathbf{DZ} - \mathbf{E}\|_F^2) \end{aligned} \tag{9}$$

where for updating $\mathbf{W}$ the constraint $\pi_\Omega(\mathbf{W}) = \mathbf{0}$ should be enforced after shrinkage.

The calculation is described in Algorithm 2

4. EXPERIMENTS

In this section, we evaluate our algorithm on two database: Extended Yale B [19] and Caltech-101 [20]. Our approach is compared with SRC [5], LC-KSVD [7], DLRD-SR [9] and Zhang's method [10]. The dictionary is initialized with randomly selected samples from each class.

4.1. Evaluation on Extended Yale B

In this subsection, we conduct experiments on Extended Yale B database, which comprises 2414 face images of 38 people, with 64 images per person and 192×168 pixels per image. This database is challenging for its varying illumination condition and expression. 1900 images, 50 per person, are randomly selected for training and the rest for testing. Each image is projected onto a 504-dimensional vector with a randomly generated matrix from zero-mean normal distribution.

We compare our approach with SRC, LC-KSVD, DLRD-SR and Zhang's method with a 570-column dictionary, 15

Algorithm 2 Low-Rank and Partial Sparse Representation Algorithm via Inexact ALM

Input:

Test sample $\mathbf{x}$; Learned dictionary $\mathbf{D}$; Training samples $\mathbf{X}^t$ and coefficient $\mathbf{Z}^t$; Parameters for Inexact ALM

Output:

$\mathbf{Z}, \mathbf{E}$

$\mathbf{X} = [\mathbf{X}^t, \mathbf{x}], \mathbf{F} = [\mathbf{Z}^t, \mathbf{0}]$

$\mathbf{J} = \mathbf{0}, \mathbf{Z} = \mathbf{0}, \mathbf{W} = \mathbf{0}, \mathbf{E} = \mathbf{0}, \mathbf{Y}_1 = \mathbf{0}, \mathbf{Y}_2 = \mathbf{0}, \mathbf{Y}_3 = \mathbf{0}$

while not converged **do**

$(\mathbf{U}, \mathbf{S}, \mathbf{V}) = \mathrm{svd}(\mathbf{Z} + \mathbf{Y}_2/\mu)$

$\mathbf{J} := \mathbf{U}\mathcal{S}_{\alpha\mu^{-1}}[\mathbf{S}]\mathbf{V}^T$

$\mathbf{W} := \mathcal{S}_{\beta\mu^{-1}}[\mathbf{Z} - \mathbf{F} + \mathbf{Y}_3/\mu]$

$\pi_\Omega(\mathbf{W}) = \mathbf{0}$

$\mathbf{Z} := (\mathbf{D}^T\mathbf{D} + 2\mathbf{I})^{-1}(\mathbf{D}^T\mathbf{X} - \mathbf{D}^T\mathbf{E} + \mathbf{J} + \mathbf{F} + \mathbf{W} + (\mathbf{D}^T\mathbf{Y}_1 - \mathbf{Y}_2)/\mu)$

$\mathbf{E} := \mathcal{S}_{\gamma\mu^{-1}}(\mathbf{X} - \mathbf{D}\mathbf{Z} + \mathbf{Y}_1/\mu))$

$\mathbf{Y}_1 := \mathbf{Y}_1 + \mu(\mathbf{X} - \mathbf{D}\mathbf{Z} - \mathbf{E})$

$\mathbf{Y}_2 := \mathbf{Y}_2 + \mu(\mathbf{Z} - \mathbf{J})$

$\mathbf{Y}_3 := \mathbf{Y}_3 + \mu(\mathbf{Z} - \mathbf{F} - \mathbf{W})$

$\mu = \min(\mu_{max}, \rho\mu)$

end while

each class. We measure the performance of SRC using two different dictionary size (15 training samples per class and all training samples per class). The comparative results are shown in Table 1. The proposed method, by taking advantage of label information of training samples in dictionary learning phase and the correlation between test sample and training samples in solving the coefficient, outperforms SRC, LC-KSVD, DLRD-SR and Zhang's method.

Table 1. Performance comparisons on the Extended Yale B

Method	Accuracy
SRC(15)	86.9%
SRC(all)	97.0%
LC-KSVD	89.1%
DLRD-SR	90.5%
Zhang's	93.8%
SC-LRPSR	**97.8%**

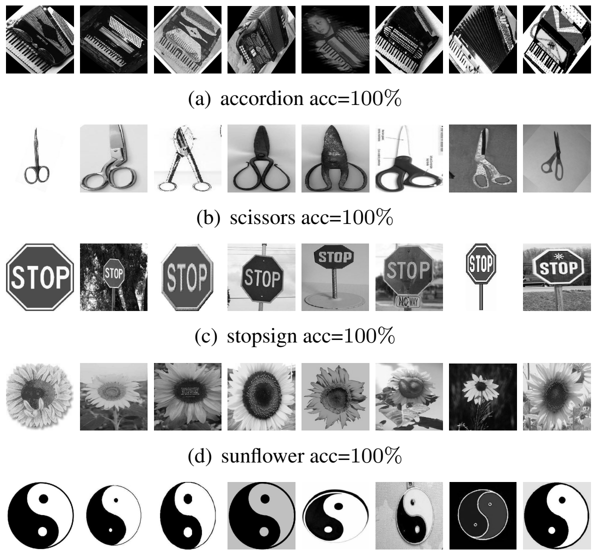
4.2. Evaluation on Caltech-101

The Caltech-101 database is consist of more than 9000 images from 102 classes (i.e. 101 object classes such as vehicles, animals, flowers, tress, etc. and a background class). The number of images in each class varies from 31 to 800. The experiment is conducted with spatial pyramid features of the images and 30 images is randomly selected as training samples every class.

We evaluate our approach and compare with SRC, LC-KSVD, DLRD-SR and Zhang's method. We measure the performance of SRC using two different dictionary size (15 training samples per class and all training samples per class). The comparative results are shown in Table 2. The proposed method outperforms SRC, LC-KSVD, DLRD-SR and Zhang's method. Figure 1 shows some examples from classes which achieve 100% classification accuracy.

Table 2. Performance comparisons on the Caltech 101

Method	Accuracy
SRC(15)	64.9%
SRC(all)	70.7%
LC-KSVD	73.6%
DLRD-SR	71.2%
Zhang's	73.6%
SC-LRPSR	**78.3%**



(a) accordion acc=100%

(b) scissors acc=100%

(c) stopsign acc=100%

(d) sunflower acc=100%

(e) yin-yang acc=100%

Fig. 1. examples from classes with 100% classification accuracy of the Caltech-101

5. CONCLUSION

In this paper, a novel Structure-Constrained Low-Rank and Partial Sparse Representation algorithm for image classification is proposed. The label information is fully exploited in dictionary learning step by imposing a structure constraint on coefficient matrix, which will also induce sparsity of coefficient matrix. Then a sparse and low-rank representation of test sample is computed via Low-Rank and Partial Sparse Representation under the assumption that the test sample is highly correlated with training samples. The experimental results confirm that our method is competitive compared with the state-of-the-art methods.

6. REFERENCES

[1] Michael Elad, Mário A.T. Figueiredo, and Yi Ma, "On the role of sparse and redundant representations in image processing," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 972–982, 2010.

[2] Michael Elad, "Sparse and redundant representation modelingłwhat next?," *Signal Processing Letters (SPL), IEEE*, vol. 19, no. 12, pp. 922–928, 2012.

[3] Yu-Wei Chao, Yi-Ren Yeh, Yu-Wen Chen, Yuh-Jye Lee, and Y-CF Wang, "Locality-constrained group sparse representation for robust face recognition," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 761–764.

[4] Angshul Majumdar and Rabab K Ward, "Fast group sparse classification," *Electrical and Computer Engineering, Canadian Journal of*, vol. 34, no. 4, pp. 136–144, 2009.

[5] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, 2009.

[6] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing (TSP), IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, 2006.

[7] Zhuolin Jiang, Zhe Lin, and Larry S Davis, "Learning a discriminative dictionary for sparse coding via label consistent k-svd," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1697–1704.

[8] Lihe Zhang, Kunyu Zhang, and Xiaoli Dong, "Online sparse learning utilizing multi-feature combination for image classification," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 197–200.

[9] Long Ma, Chunheng Wang, Baihua Xiao, and Wen Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2586–2593.

[10] Yangmuzi Zhang, Zhuolin Jiang, and Larry S Davis, "Learning structured low-rank representations for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 676–683.

[11] Lihe Zhang and Chen Ma, "Low-rank, sparse matrix decomposition and group sparse coding for image classification," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 669–672.

[12] Mingyang Jiang and Jufu Feng, "Robust low-rank subspace recovery and face image denoising for face recognition," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 3033–3036.

[13] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11:1–11:37, 2011.

[14] Emmanuel J Candès and Benjamin Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[15] Ehsan Elhamifar and René Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on*, vol. 35, no. 11, pp. 2765–2781, 2012.

[16] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma, "Robust recovery of subspace structures by low-rank representation," *Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on*, vol. 35, no. 1, pp. 171–184, 2012.

[17] Guangcan Liu, Zhouchen Lin, and Yong Yu, "Robust subspace segmentation by low-rank representation," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 663–670.

[18] Zhouchen Lin, Minming Chen, Leqin Wu, and Yi Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," Tech. Rep. UILU-ENG-09-2215, UIUC, 2009.

[19] 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," *Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on*, vol. 23, no. 6, pp. 643–660, 2001.

[20] L. Fei-Fei, R. Fergus, and Pietro Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," in *IEEE CVPR Workshop of Generative Model Based Vision (WGMBV)*, 2004.