# Robust Face Recognition with Locality-Sensitive Sparsity and Group Sparsity Constraints

Xi Sun, Wentao Chan, and Lei Qu(✉)

Key Laboratory of Intelligent Computing and Signal Processing,
Anhui University, Hefei 230039, China
qulei@ahu.edu.cn

**Abstract.** In this paper, we present a robust face recognition method with combined locality-sensitive sparsity and group sparsity constraint. The group sparsity constraint is designed to utilize the grouped structure information embedded in the training data. Its key idea is to try representing the test image with training images from fewer individuals. We show that, by further integrating the local similarity information between the test image and training images, the embedded group structure information can be better utilized, and as result, the recognition performance can be significantly improved. Experimental results on the ORL, AR and Extended Yale B database verify the superiority of our proposed method under different pose, illumination, expression variations and different dimension reduction settings.

**Keywords:** Face recognition · Sparse representation · Locality-sensitive · Group sparsity

## 1 Introduction

Face recognition is one of the most challenging research topics in the field of pattern recognition and computer vision. It has many applications in business and law enforcement besides its scientific significance, for instance, supervision, security, communication, human-computer interaction, etc. After 30 years of research, Numerous face recognition methods have been proposed by researchers and scientists. However, robust face recognition is still an open problem due to the complexity arising from expressions, hairstyle, post, illumination variations and the similarity of the facial organ distribution among different peoples.

In the past few years, due to the rising of Compressive Sensing, especially the core technology of sparse representation which can not only reduce the cost of data analysis and processing, but also improve the efficiency of data compression. Method based on sparse representation received extensive attention because of its excellent performance and robustness to noise and occlusion. In 2009, Wright et. al. [1] introduced sparse representation to solve the face recognition problem successfully and proposed Sparse Representation Classification (SRC) method. SRC looks for the sparsest representation of a test sample in the dictionary which composed of all training data, and can be


Y.-J. Zhang (Ed.): ICIG 2015, Part III, LNCS 9219, pp. 260–270, 2015.
DOI: 10.1007/978-3-319-21969-1_23

solved efficiently via $l_1$-norm minimization. SRC is effective and robust in the classification, but the interior structure information of the training data dictionary has not been considered. Yang and Zhang [2] proposed a Gabor-feature based SRC (GSRC) scheme in 2010, which uses the image local Gabor features for SRC. The significant superiority of GSRC is its compact occlusion dictionary, which can greatly reduce the computational cost of sparse coding. Although GSRC has much better performance than SRC, the interior structure information has not been taking into account either. Yuan et al. [3] proposed the group lasso which could solve the convex optimization problem at the group level. In 2011, Elhamifar et al. [4] proposed a more robust classification method that using group sparse representation, which searching a representation that uses the minimum number of blocks. It overcomes the drawback of SRC, however, group sparse representation loses to capture the locality structure of data. Friedman et al. [5, 6] proposed an idea of "A sparse-group lasso", which includes both individual and group sparse representation classification. When the groups consist of linearly independent data, this method can well solve face recognition problem. In recent years some initial efforts have been devoted to develop kernel sparse representation classification (KSRC) [7, 8]. It integrates the kernel method and the SRC method, so that KSRC has been successfully applied for image classification and face recognition, but it is not able to capture the locality structure information. Shrivastava et al. [9] proposed a multiple kernel learning (MKL) algorithm, which is based on the SRC method. It makes full use of the nonlinear kernel SRC in efficiently representing the nonlinearities in the high dimensional feature space, but still could not obtain the locality structure of data. In 2014, Zhang and Zhao [10] integrate KSRC with data locality in the kernel feature space, and develop an extension of KSRC, called locality-sensitive kernel sparse representation classification (LS-KSRC). Due to integrating the data locality, it can preserve the similarity between the test sample and its neighboring training data when searching sparse representation coefficients. This is good property for classification.

In this paper, inspired by the cogitation of LS-KSRC [10], we introduce a new method which combined the locality-sensitive sparsity with the group sparsity constraint. It not only takes account of the grouped structure information of the training data dictionary, but also integrates the data locality. The proposed method aims to learn both group sparsity and data locality at the same time, and achieve improved classification.

The rest of this paper is organized as follows. In Sect. 2, we review the sparse representation classification method and the group sparsity classification method in brief. Section 3 provides the method that we proposed in detail. Experimental results and analysis are presented in Sect. 4. Section 5 conclusions this paper with a summary.

## 2 Sparse Representation

### 2.1 Sparse Representation-Based Classification

The SRC method can be seen as a process that converts the input test image into the sparsest linear combination of training images with illumination, expression, etc.

variations. Suppose we have $n$ classes, the *ith* class has $n_i$ training samples, $a_{i,j} \in R^{D\times 1}$ as the image feature vector of the *jth* image in the *ith* class, $D$ denotes dimension of the image feature vector, $A_i = [a_{i,1}, a_{i,2}, \ldots, a_{i,n_i}] \in R^{D\times n_i}$ contains training images of the *ith* class. Let $A = [A_1, A_2, \ldots, A_n]$ be the entire training set, $y$ denotes an input test image. To avoid the NP-hard problem result from the $l_0$-norm, practically, the SRC often refers to solving the following $l_1$-norm minimization problem,

$$\min_{x\in R^D} \|x\|_1 \quad s.t. \quad \|y - Ax\|_2 < \varepsilon, \tag{1}$$

where $x$ is the sparse coefficient vector, $\varepsilon$ is associated with a noise term with bounded energy. The optimal solution is denoted by $x^{*T} = [x_1^{*T}, x_2^{*T}, \ldots, x_n^{*T}]$, $x_i$ contains the coefficients associated with the *ith* class.

## 2.2 Group Sparsity Classification

Although great success can be obtained by the SRC, one potential problem of this method is that the test image may be represented by the training images from different individuals. For the task of face recognition, this problem may result in ambiguous or even wrong recognition. Ideally, the test image should be only represented by the training images from only one individual which corresponding to the correct classification. Based on this idea, Elhamifar et al. proposed a more robust grouped sparse representation based classification method [4] which tries to represent the test image by the training images from as fewer individuals as possible. To implement this constraint, the training data dictionary is divided into groups where each group is form by the training images from the same individual. Then, the recognition is realized by searching a representation that uses the minimum number of group. This is equivalent to convert the classification problem to a structural sparse recovery problem.

Given a test image $y$, the following convex problem is considered to derive the sparse coefficient $x$ of $y$, which minimizes the number of nonzero groups $x_i$ from the dictionary,

$$P_1: \quad \min_{x\in R^D} \sum_{i=1}^{n} \|x_i\|_2 \quad s.t. \quad \|y - Ax\|_2 < \varepsilon, \tag{2}$$

where $x_i$ represents the coefficients associated with the *ith* class, $\varepsilon$ is associated with a noise term with bounded energy, $A = [A_1, A_2, \ldots, A_n]$ is denoted as the entire training set.

In addition to minimize the number of nonzero groups, one alternative method is to minimizes the number of nonzero reconstructed vectors $A_i x_i$,

$$P_2: \quad \min_{x\in R^D} \sum_{i=1}^{n} \|A_i x_i\|_2 \quad s.t. \quad \|y - Ax\|_2 < \varepsilon, \tag{3}$$

where $A_i = [a_{i,1}, a_{i,2}, \ldots, a_{i,n_i}] \in R^{D\times n_i}$ denotes the *ith* training sample.

From the Eqs. (2) and (3), we note that they are equivalent only if the groups consist of linearly independent data. Because of the similarity of the facial organ distribution among different people, it is very easy to appear linearly dependent data in face recognition application.

## 3 The Proposed Method

Nowadays in many pattern recognition problems, data locality has been widely used, such as K-nearest neighbor (KNN) classifier [11], data clustering [12], and image classification [13] et al. And it has been pointed out in [14], which data locality is more essential than sparsity for sparse coding. As mentioned in the introduction, group sparse representation loses to capture the locality structure of data. To overcome this drawback, now we propose a more advanced method which combined the locality-sensitive sparsity with the group sparsity constraint.

First, we consider one type of group sparse representation in our method, which minimizes the number of nonzero groups from the dictionary. By means of enforcing data locality in the kernel feature space to the $l_1$-norm minimization problem in LS-KSRC [10], the $l_1$-norm minimization problem in our proposed method is formulated as follows,

$$P3: \quad \min_{x \in R^D} \lambda \|p \bullet x\|_2 + \beta \sum_{i=1}^{n} \|x_i\|_2 \quad s.t. \quad \|y - Ax\|_2 < \varepsilon \tag{4}$$

where $\lambda$ is the regularization parameter, the symbol $\bullet$ represents element-wise multiplication, $x$ denotes the sparse coefficient vector, $\beta$ weights the group sparsity regularizer, $A = [A_1, A_2, \ldots, A_n]$ is the concatenation of training samples from all the classes, $y$ denotes a test sample, $\varepsilon$ is associated with a noise term with bounded energy, $x_i$ is the representation coefficient associated with the *ith* training sample. $p$ is the locality adaptor, and we use the following exponential locality adaptor like the LS-KSRC [10] in our method,

$$p = \sqrt{\exp\left(\frac{d(y_i, y_j)}{\eta}\right)} \tag{5}$$

where $\eta$ is a positive constant, $y_i$ denotes a test sample, $y_j$ denotes a neighboring training sample of $y_i$, and $d(y_i, y_j)$ induced by a $l_2$-norm is defined as,

$$d(y_i, y_j) = \|y_i - y_j\|_2 \tag{6}$$

where $d(y_i, y_j)$ denotes the Euclidean distance.

In Eq. (4), the vector $p$ is used to measure the distance between a test sample and each column of training sample. In another word, the vector $p$ can be seen as a dissimilarity vector, and is applied to constraint the corresponding sparse coefficient. It should be noted that since the solutions only have few significant values whereas most

coefficients are zero, the resulting coefficients of Eq. (4) is regarded to be still sparse in the sense of $l_1$-norm, but not sparse in the sense of $l_2$-norm. Thus minimizing the problem in Eq. (4) means to encode the test sample with its neighboring training samples, and make the proposed method integrates both group sparsity and data locality structure while obtaining the optimal sparse coefficients. This guarantees our method with a good ability of learning discriminating sparse representation coefficients for classification.

Then, similarly, we consider another way to optimize the group sparse representation in our proposed method, which minimizes the number of nonzero reconstructed vectors. And the $l_1$-norm minimization problem in our proposed method is formulated as follows,

$$P4: \quad \min_{x \in R^D} \lambda \|p \bullet x\|_2 + \beta \sum_{i=1}^{n} \|A_i x_i\|_2 \quad s.t. \quad \|y - Ax\|_2 < \varepsilon \tag{7}$$

where $A_i \in R^{D \times n_i}$ is the subset of the training samples from class $i$.

It is worth pointing out that in Eqs. (4) and (7), the first term constraints the data locality, it can preserve the similarity between the test sample and its neighboring training data, and give sparse representation coefficients with discriminating information. The second term constraints group sparsity, it takes account of the grouped structure information embedded in the training data. As a result, our method integrates group sparsity constraints and data locality structure at the same time. Our method which combines locality-sensitive sparsity with group sparsity constraints is summarized in Algorithm 1.

**Algorithm 1.** The proposed method

(1) Input: the matrix of all training samples $A$, and a test sample $y$

(2) Calculate the dissimilarity vector between a test sample and each training sample by using the exponential locality adaptor $p$

(3) Solve the $l_1$-norm minimization problem

(4) Compute the residuals by using the samples associated with the $ith$ class by

$$r_i(y) = \min_{i=1,2..n} \left( \left\| y - A_i x_i^{*T} \right\|_2 \right)$$

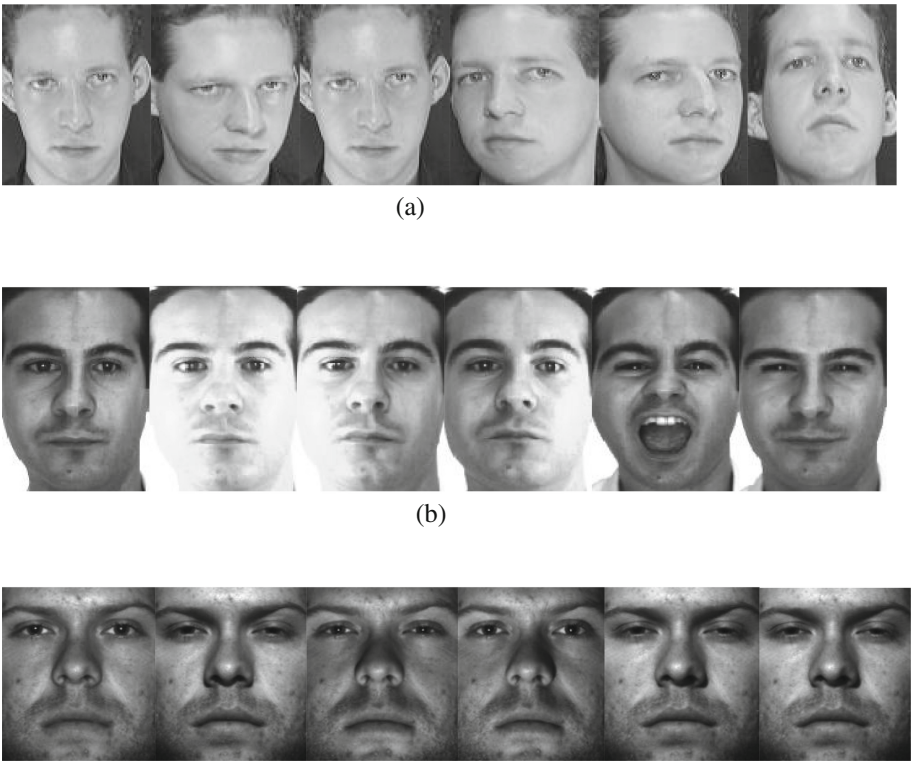
(5) Output: the class label $y$ of the given test sample

$$class(y) = \arg\min_{i=1,2..n} r_i(y)$$

where $x_i^{*T} = [x_1^{*T}, x_2^{*T}, \ldots, x_n^{*T}]$ denotes the optimal solution, which contains the coefficients associated with the *ith* class.

## 4 Experiments

In this section, we evaluate the performance of our proposed method on several commonly used face databases, including ORL [15], AR [16], and the Extended Yale B [17]. Figure 1 shows some sample images from three benchmarking face databases, where various pose, expression and illumination variations can be observed. To illustrate the superiority of our method, the recognition performance of several closely related sparse representation based face recognition methods [1, 4, 10] are investigated and compared.



(a)

(b)

(c)

**Fig. 1.** Sample images from three face databases: (a) the ORL database, (b) the AR database and (c) the extended Yale B database.

For all experiment, the Principal Component Analysis (PCA) [18] is used to reduce the feature dimension before classification is performed, and the CVX toolbox is employed to solve the $l_1$-norm minimization problem. We choose $\varepsilon$=0.05, $\eta$ = 0.5 for all experiments. Since for different methods and databases, the best performance was achieved with different $\lambda$ and $\beta$ settings. To conduct a fair comparison, we tested different $\lambda$ and $\beta$ combination for all methods, only their best performances were recorded and compared.

### 4.1 Experiments on ORL Database

In this experiment, we tested the performance of different methods under different training/test split settings on the ORL database. The ORL database contains 400 face images from 40 subjects (10 different images per subject) with variations in poses, illuminations and facial expressions. For each test, we randomly select a subset with L (L = 3, 4, 5) images per subject to form the training set, and the rest images was taken as testing set. The recognition accuracy is calculated as the average recognition rate of 30 random tests. Table 1 gives the recognition rate of different methods.

**Table 1.** The performance comparisons on the ORL database.

| L（PCA） / method | 3（40） | 4（40） | 5（40） |
|---|---|---|---|
| SRC | 86.286% | 90.000% | 93.500% |
| LS-KSRC | 89.643% | 92.750% | 95.500% |
| P1 | 87.857% | 89.167% | 93.000% |
| P2 | 89.286% | 92.917% | 96.500% |
| P3 | 90.357% | 94.167% | 97.000% |
| P4 | 94.286% | 95.417% | 98.000% |

As shown in Table 1, we can see that both of our method (P3 and P4) outperform the other used methods under different training/test dataset split settings, including SRC, LS-KSRC and two types of group sparsity. P3 method obtains the best recognition performance with an accuracy of 90.357 % for 3 Train, 94.167 % for 4 Train, and 97.000 % for 5 Train, and our P4 method obtains the best recognition performance with an accuracy of 94.286 % for 3 Train, 95.417 % for 4 Train, and 98.000 % for 5 Train. And we can also note that in each setting, the highest recognition accuracy was always obtained by the P4 method. For instance, in the case of L = 3, the performance of P4 method is about 8 % better than SRC, 4.643 % better than LS-KSRC, 6.429 % better than P1, 5 % better than P2, and even 3.929 % better than our P3 method. These indicate that our method is a more effective classification method for face recognition in comparison with the other used methods.

### 4.2 Experiments on AR Database

In this experiment, the performance of different methods was evaluated on the AR database under different PCA feature dimension reduction settings. The AR database contains 3276 face images from 126 subjects with various poses, expressions, and illuminations. The original size of images is 165 * 120. We choose a subset of 1400 images from 100 subjects (50 male and 50 female), to ensure that for each subject, there are 14 frontal face images. In each test, we randomly select 7 images per subject to form the training dataset, and the rest is use as test dataset. Same as the experiments on the ORL database, the recognition accuracy is calculated as the average recognition rate of 30 random tests. Table 2 gives the best recognition rate of different methods under different PCA feature dimension reduction settings.

**Table 2.** The performance comparisons on the AR database.

| reduced dimension / method | 36 | 54 | 130 |
|---|---|---|---|
| SRC | 73.104% | 79.828% | 86.266% |
| LS-KSRC | 73.247% | 80.687% | 87.554% |
| P1 | 72.818% | 81.402% | 86.981% |
| P2 | 74.678% | 81.545% | 88.698% |
| P3 | 75.393% | 81.688% | 87.838% |
| P4 | 78.112% | 82.546% | 89.557% |

As shown in Table 2, both our method (P3 and P4) show their superiority over other methods under different PCA dimension reduction settings. The highest recognition accuracy obtained by P3 is 75.393 % for 36 dimension, 81.688 % for 54 dimension, and 87.838 % for 130 dimension. Our P4 method shows better recognition performance than the P3 with an accuracy of 78.112 % for 36 dimension, 82.546 % for 54 dimension, and 89.557 % for 130 dimension. These demonstrate the advantage of our method as a classifier for face recognition again, especially in the low dimension

cases. When the reduced dimension is 36, the P3 method can achieve 2.289 % improvement over SRC, 2.146 % over LS-KSRC, and 2.575 % over P1, the P4 method can achieve 5.008 % improvement over SRC, 4.865 % over LS-KSRC, and 3.434 % over P2.

### 4.3 Experiments on Extended Yale B Database

The Extended Yale B database is composed of 2414 cropped frontal face images from 38 subjects. For each subject, there are about 64 face images of size 192 * 168 with different illuminations. We select a subset with 32 images per individual for training, and the rest images are used for testing. Table 3 gives the best recognition accuracy of different methods under different PCA dimension reduction settings.

**Table 3.** The performance comparisons on the extended Yale B database.

| reduced dimension / method | 36 | 54 | 130 |
|---|---|---|---|
| SRC | 87.572% | 89.877% | 94.897% |
| LS-KSRC | 88.230% | 89.959% | 95.073% |
| P1 | 80.412% | 90.123% | 94.815% |
| P2 | 82.469% | 92.827% | 95.556% |
| P3 | 89.043% | 91.275% | 95.274% |
| P4 | 90.041% | 93.004% | 96.626% |

As can be seen from the results in Table 3, the proposed methods (P3 and P4) still outperform others methods. The P3 method obtains the best recognition performance with an accuracy of 89.043 % for 36 dimension, 91.275 % for 54 dimension, and 95.274 % for 130 dimension. In the low dimension cases, our method shows much better performance than other methods. This is consistent with the previous experimental results on the AR database. For instance, when the feature dimension is reduced

to 36, the P4 method obtains the best recognition performance with an accuracy of 90.041%. This performance is about 2.469 % better than SRC, 1.811 % better than LS-KSRC, 9.629 % better than P1, and 7.572 % better than P2. The P4 method obtains the highest performance (96.626 %) when the feature dimension is setting to 130.

## 5 Conclusions

In this paper, we propose a novel classification method with locality-sensitive sparsity and group sparsity constraints for robust face recognition. This method learns group sparsity and data locality at the same time. It not only takes into account the grouped structure information of the training data dictionary, but also integrates the data locality, thus can learn more discriminating sparse representation coefficients for face recognition. To testify the effectiveness of the method that we proposed, we perform experiments for face recognition on the ORL, AR, and Extended Yale B databases and demonstrate the power of our algorithm compared to some other methods. Experimental results show that our method achieves very promising recognition results on these datasets.

**Acknowledgements.** This work was supported by Chinese Natural Science Foundation Project (No. 61201396, 61301296, 61377006, U1201255); Anhui Provincial Natural Science Foundation (No. 1508085MF120); Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry; Technology Foundation for Selected Overseas Chinese Scholar, Ministry of Personnel of China.

## References

1. Wright, J., Yang, A.Y., Ganesh, A., et al.: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 210–227 (2009)
2. Yang, M., Zhang, L.: Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 448–461. Springer, Heidelberg (2010)
3. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped varables. J. Roy. Stat. Soc. B (Stat. Methodol.) **68**(1), 49–67 (2006)
4. Elhamifar, E., Vidal, R.: Robust classification using structured sparse representation. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1873–1879. IEEE (2011)
5. Friedman, J., Hastie, T., Tibshirani, R.: A note on the group lasso and a sparse group lasso. arXiv preprint arXiv:1001.0736 (2010)
6. Simon, N., Friedman, J., Hastie, T., et al.: A sparse-group lasso. J. Comput. Graph. Stat. **22** (2), 231–245 (2013)
7. Zhang, L., Zhou, W.-D., Chang, P.-C., Liu, J., Yan, Z., Wang, T., Li, F.-Z.: Kernel sparse representation-based classifier. IEEE Trans. Signal Process. **60**, 1684–1695 (2012)
8. Zhou, Y., Gao, J., Barner, K.E.: An enhanced sparse representation strategy for signal classification. In: SPIE 8365, Compressive Sensing, Baltimore, p. 83650H (2012)

9. Shrivastava, A., Patel, V., Chellappa, R.: Multiple kernel learning for sparse representation-based classification. IEEE Trans. Image Process. **23**(7), 3013–3024 (2014)
10. Zhang, S., Zhao, X.: Locality-sensitive kernel sparse representation classification for face recognition. J. Vis. Commun. Image Represent. **25**, 1878–1885 (2014)
11. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Trans. Inf. Theor. **13**, 21–27 (1967)
12. Cai, D., He, X., Han, J.: Document clustering using locality preserving indexing. IEEE Trans. Knowl. Data Eng. **17**, 1624–1637 (2005)
13. Wang, J., Yang, J., Yu, K., Lv, F., et al.: Locality-constrained linear coding for image classification. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010), San Francisco, pp. 3360–3367 (2010)
14. Yu, K., Zhang, T., Gong, Y.: Nonlinear learning using local coordinate coding. Adv. Neural Inf. Process. Syst. **22**, 2223–2231 (2009)
15. Samaria, F.S., Harter, A.C.: Parameterisation of a stochastic model for human face identification. In: The Second IEEE Workshop on Applications of Computer Vision, Sarasota, pp. 138–142 (1994)
16. Martinez, A.M.: The AR face database. CVC technical report, p. 24 (1998)
17. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 643–660 (2001)
18. Jolliffe, I.T.: Principal Component Analysis. Springer, New York (1986)